THE TODDLER REMOTE ASSESSMENT OF VIRTUAL EYE TRACKING AND LANGUAGE (TRAVEL) STUDY: AN EYE TRACKING AND BEHAVIORAL ASSESSMENT OF JOINT ATTENTION AT HOME

Jessica Elizabeth Goldblum

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements of Doctor of Philosophy in the Department of Psychology and Neuroscience in the College of Arts and Sciences (Developmental Psychology).

Chapel Hill
2023

Approved by:

Cathi B. Propper

Clare E. Harrop

Jessica Cohen

Shauna Cooper

Kristin Lindquist

ABSTRACT

Jessica Elizabeth Goldblum: The Toddler Remote Assessment of Virtual Eye Tracking and Language (TRAVEL) Study: An Eye Tracking and Behavioral Study of Joint Attention at Home
(Under the direction of Cathi Propper and Clare Harrop)

The current study tested the feasibility, utility, and validity of the Toddler Remote Assessment of Virtual Eye Tracking and Language (TRAVEL) Study, a remote study bringing portable eye tracking and behavioral measures of joint attention (JA) home to families. Joint Attention (JA) is the ability to coordinate the attention of others to an object, entity, or event and difficulties with JA can be seen in clinical populations, most commonly autism. Developing home-based, research-grade measures of JA with high clinical utility could serve as a supplement to diagnostic and intervention tools for autism and substantially diversify research populations.

Fifty children (25 who were autistic and 25, non-autistic) and their caregivers completed a battery of experiential eye tracking, behavioral, and caregiver-reported tasks of JA at home in addition to a feasibility survey. Logistic regression was leveraged to estimate children's gaze trajectories of JA (*TRAVEL trajectories*) in response to specific events while exploratory factor analysis was used to create global estimates of children's JA (*Total TRAVEL scores*) on eye tracking and behavioral measures. We hypothesized that (1) TRAVEL would be a feasible means of remote assessment and that TRAVEL outcomes would (2) discriminate between groups and (3) exhibit construct validity in comparison to caregiver-reported child JA, language, and restricted and repetitive behaviors (RRBs). Caregivers reported that TRAVEL was a largely feasible means of remote assessment with a few areas of improvement noted. Several TRAVEL outcomes significantly discriminated between groups and *Total TRAVEL scores* had good clinical utility. TRAVEL outcomes did not exhibit adequate construct validity with caregiver-reported measures of JA, language, or RRBs but more research is needed with changes to some survey and paradigm measures. Overall, the TRAVEL Study has important implications for future pediatric research by bringing research-grade eye tracking tools home to families.

ACKNOWLEDGEMENTS

There are a number of individuals I would like to thank for making this work possible. Thank you to the families who participated in the TRAVEL Study. You showed up for your kids. Thank you to my primary and secondary graduate advisors, Drs. Cathi Propper and Clare Harrop. Thank you for the opportunity to be your mentee and to learn and grow under your wings. I am grateful to have been advised by two strong female researchers whose achievements I can only hope to emulate in my own career. I also would not have made it through graduate school without my Slack confidant, Amanda Wylie. Thank you for always being there and understanding grad school when no one else did. Finally, I am gracious for my respective labs – the BEE Lab and Harrop Lab – and all of the wonderful researchers, grad students, post-docs, and staff I was fortunate enough to work with during my time at UNC.

None of this would have been possible without my family and friends. To my parents and my sister, Zoe, for your unwavering love and support even in the face of life's biggest challenges. You made me into the person I am today and for that I am forever grateful. To my aunt Margie, uncle Perry, my grandparents, and the rest of my family in New Mexico, thank you for always supporting and encouraging me from afar. Thank you to my friends from high school (Gabi), Emory (Jennie, Linda, Lauren, Samiha, and Grace), and UCLA and LA for keeping me sane and bringing me joy even if we only saw each other every so often the past few years. Thank you to my second family, Linda, Nathan, Tova, and Shayna, for loving me like your own daughter/sister. Most importantly, to my person, Jake. You have been there for me through thick and thin, in the best and worst of times. I truly could not have done this without you. Thank you

for being the father to my furballs (Junior and Tallulah) and my best friend. I love you more than words can say and cannot wait to become your wife in a little over a month.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAN	Anterior Attention Network
ADOS-2	Autism Diagnostic Observation Schedule – Second Edition
ANOVA	Analysis of Variance
AUC	Area Under the Curve
BOSA	Brief Observation of Symptoms of Autism
CFI	Comparative Fit Index
CI	Confidence Interval
EFA	Exploratory Factor Analysis
ELA	Elevated likelihood for an autism diagnosis
IJA	Initiating Joint Attention
IRR	Inter-rater Reliability
JA	Joint Attention
LCA	Latent Curve Analysis
MCDI	MacArthur Bates Communicative Development Inventory
MCDI-EV	MacArthur Bates Communicative Development Inventory Expressive Vocabulary Score
MCDI-RV	MacArthur Bates Communicative Development Inventory Receptive Vocabulary Score
ML	Maximum Likelihood
MSD	Motor and Social Development Scale
NODA	Naturalistic Observation Diagnostic Assessment
PAN	Posterior Attention Network
PCX	Parent-Child Interaction

PDPM	Parallel and Distributed Information-Processing Model
RA	Research assistant
RMSEA	Root Mean Square Error of Approximation
RJA	Responding to Joint Attention
ROC	Receiver Operator Characteristic
RTN	Response to Name
SCQ	Social Communication Questionnaire
SD	Standard Deviation
SEM	Structural Equation Modeling
SORF	Systematic Observation of Red Flags
SRS-2	Social Responsiveness Scale – Second Edition
SRS-RRBs	Social Responsiveness Scale – Restricted, Repetitive Behaviors and Interests Factor Score
TAP	TELE-ASD-PEDS
TLI	Tucker-Lewis Index
TPVT	NIH Toolbox Picture Vocabulary Test
TRAVEL	Toddler Remote Assessment of Virtual Eye Tracking and Language

CHAPTER 1: INTRODUCTION

1.1 Study Overview

Joint attention (JA) is the innately human use of attention and language to share information with others – a form of social communication – that develops in the first and second years of life (Mundy & Newell, 2007). JA assessment in infancy and early childhood is important across populations because JA is foundational for social cognition and lifelong mental health (Mundy & Sigman, 2015). JA in neurotypical development is predictive of cognition, social competence, behavioral regulation, and language ability throughout childhood (Acra et al., 2009; Sheinkopf et al., 2004; Vaughan Van Hecke et al., 2012). Many neurodevelopmental batteries assess JA because it can be affected in clinical populations including children with attention deficit/hyperactivity disorder (ADHD) (Baird et al., 2000), rare neurogenetic syndromes such as Prader-Willi Syndrome (Lin et al., 2003), and Down Syndrome (Kasari et al., 1995). JA is most often assessed in the context of autism. Autistic children show delays and challenges with JA compared to non-autistic children such that phenotypic differences in JA are a core feature and diagnostic indicator of autism (Lord et al., 2000; Loveland & Landry, 1986).

Many children experience barriers that impede opportunities for developmental screening, neurodevelopment assessments, and to participate in research. There exists an immense shortage of behavioral health providers in the United States and waitlists for neurodevelopmental assessment can be staggering (Counts, 2023). Training on many neurodevelopmental batteries that assess JA, such as the Autism Diagnostic Observation

Schedule (ADOS-2; Lord et al., 2000) can be time intensive, expensive, and exclusive to experienced professionals or clinicians and further limits the pool of available providers. Simply traveling to a clinic can be burdensome for individuals living in geographically diverse areas or who may not have the means or resources to acquire time off work, childcare, or travel expenses, which can limit diversity in research populations. Barriers to programs that assess JA are particularly problematic for autistic populations because JA is a diagnostic indicator of autism and obtaining an official autism diagnosis is essential for obtaining therapeutic services that improve linguistic, cognitive, and behavioral outcomes when started early (Estes et al., 2015; Vinen et al., 2022). JA assessment in early childhood can therefore be a "gatekeeper" towards diagnosis and subsequent effective treatment for autistic children.

Remote neurodevelopmental assessment is one possible solution to this problem as it could bring JA assessment tools home to families. However, the sudden need for such tools during the COVID-19 pandemic highlighted just how few efficacious remote neurodevelopmental programs actually exist (Ellison et al., 2021). The few remote programs that *do* exist have several notable limitations. Existing remote tools are largely behaviorally based. For example, several remote autism assessments rely on caregivers to lead their children through behavioral paradigms (while simultaneously videorecording) and videos of the assessment are later rated by clinicians reliable on the ADOS-2 (Dow et al., 2021; Nazneen et al., 2015; Wagner et al., 2021). This is in direct contrast to in-person assessments which rely on behavioral *and* technological methods to acutely assess and monitor children and make the best possible diagnoses. It is problematic that remote neurodevelopmental programs are based solely on behavioral measurement because more sensitive tools are necessary to evaluate fine grained differences in processes such as JA. Furthermore, to our knowledge there is no remote neurodevelopmental battery that has used the most sensitive measure available to assess human attention: eye tracking.

Eye tracking has become the gold standard measurement tool for attention because it is ideal for use with children. Eye tracking is non-invasive, accounts for variables such as head movement, and provides indicators across multiple levels from precise gaze location to pupillary dilation (Sasson & Elison, 2012). We are unaware of any remote neurodevelopmental assessment batteries that utilize eye tracking despite this. This is likely because eye trackers are typically manufactured as stationary devices for clinic use. Portable eye trackers are a novel concept altogether and only a handful have been developed that capture research-grade data. Moreover, portable eye trackers are still typically used with an *in-person* examiner out in the field. We are aware of only one study that has used portable eye tracking to examine JA in children *remotely*. Benson-Goldberg & Erickson (2021) used a portable eye tracker remotely for a case study to assess a child's responsiveness to a home-based emergent literacy reading intervention and found that remote eye tracking was a feasible method for acquiring JA data collected by a caregiver at-home. Remote portable eye tracking is a critical next step in the field of remote neurodevelopmental assessment because it has the potential to bring research-grade measures of JA home to families.

This study pilots the feasibility, utility, and validity of the Toddler Remote Assessment of Virtual Eye tracking and Language (TRAVEL) Study in sample of autistic and non-autistic children. TRAVEL is a novel, multimodal and remote assessment program that uses portable research-grade eye tracking together with behavioral and caregiver-report paradigms to measure JA at home. We hope to develop TRAVEL as a remote assessment that mimics lab-based JA eye tracking assessments in order to bring eye tracking research home to families. TRAVEL could also be used as a remote supplement to in-person diagnostic assessments for autism and to monitor treatment progress for JA-based interventions since JA assessment is critical for autism diagnosis and treatment. The aims of this study are threefold. First, we aim to determine if TRAVEL is a feasible method with which to remotely measure JA using eye tracking and behavioral paradigms. Second, we aim to test whether TRAVEL outcomes can differentiate autistic children from non-autistic children. Third, we aim to examine the construct validity of TRAVEL outcomes.

Current remote neurodevelopmental assessments rely on behavioral measurement, require trained clinicians for scoring, and take at least one hour to complete (Dow et al., 2020, 2021). TRAVEL introduces a battery of home-based remote portable eye tracking and behavioral paradigms that systematically and accurately measure JA in children in about thirty minutes. TRAVEL has the potential to change pediatric eye tracking research altogether. Eye tracking research has historically been conducted in labs or clinics with largely homogenous samples. TRAVEL could initiate home-based pediatric eye tracking research and would bring accessible, research-grade technology home to families of diverse backgrounds and contexts who may not otherwise have the opportunity to participate in research.

1.2 Joint Attention and Bases for Non-Autistic Development

The first and second year of non-autistic development are marked by the young child's extraordinary ability to respond to and coordinate social signals of engagement (i.e., "shareable moments") with other people through the early neurodevelopmental process of JA. JA defines the shared nature of the human experience by allowing even young infants to engage in early and meaningful social exchanges with others (Bakeman & Adamson, 1984). JA is one of the earliest forms of social learning in which infants begin to attend to and make meaning of their own social attention and that of others (Corkum & Moore, 1998). Over time and with experience, infants begin to initiate JA by conveying social information through social communicative means (e.g., eye contact, verbal and non-verbal language, affectivity) and understand that these behaviors elicit social responses from others (Salley et al., 2016). JA allows infants to participate in, reciprocate, and learn from countless social interactions that ultimately shape effective social communication. Effective social communication sets the stage for language learning (Bruner, 1974), emotion regulation (Morales et al., 2005), and social cognition and competence in early childhood (Acra et al., 2009; Mundy & Newell, 2007; Sigman & Ruskin, 1999; Vaughan Van Hecke et al., 2007). The early development of JA is overall vital for the capacity to form social relationships and for lifelong health and well-being.

The **P**arallel and **D**istributed Information-**P**rocessing **M**odel (PDPM) of JA (Mundy, 2013, 2017; Mundy et al., 2009) defines the early neural organization of JA according to an information-processing system that develops during episodes of coordinated attention with others. Infants integrate information from their own self-perception with information they perceive about the social attention of others. Infants then unify this information with internal representations that ultimately form symbolic thought in the second year of life (Mundy et al., 2009). The concurrent unification of self-perception, social coordination, and attention with internal representations is a cause and consequence of a functionally connected and distributed neural network (Mundy et al., 2009). This distributed neural network serves as a “social executive function” system that, with practice, inhibits automatic attentional behavior in favor of goal directed, volitional, and planned attentional behavior that is focused on others (Mundy & Jarrold, 2010). The information processing that takes place via this “social executive function” system continues to be developed and used throughout life, supporting deeper levels of cognition, language, and social learning (Mundy et al., 2009). We hereafter review how this distributed cortical network manifests from inception.

Non-autistic infants show a robust visual preference for social stimuli after the first days of birth (Bardi et al., 2011; Simion et al., 2008) and acquire the capacity for social perception and processing in the first three months of development with the onset of active vision (Canfield & Kirkham, 2001; Posner & Rothbart, 1998). Non-autistic infants correspondingly develop the ability to respond to joint attention (RJA), or to *follow, respond to*, and *understand* the social signals of others between nine and ten months of age (Mundy et al., 2003). RJA begins in the first three months of life with the development of gaze following (Brooks & Meltzoff, 2005). Gaze following is thought to be an automatic process in which the posterior attention system (PAN) controls attention orienting to relevant social stimuli within infants' external environments (Mundy & Vaughan Van Hecke, 2008). During episodes of social engagement when caregivers identify an object, entity, or event, infants use gaze following to detect the appropriate referent (Baldwin, 1995; Mundy & Newell, 2007). Infants ultimately develop true RJA which combines gaze following and social perception with the ability to mentalize or *understand* the social signals one is perceiving and following (Bedford et al., 2012). When infants orient to the social information (or JA) of others, the PAN iteratively processes these social indicators (e.g., another's eye and head orientation), constructs internal cognitive representations of this stimuli, and ultimately gives infants the capacity to follow, understand, and respond (Mundy & Jarrold, 2010).

The PAN is not specific to humans in that several primates and canines have the capacity to perceive and respond to social signals of engagement (Emery, 2000; Mundy & Jarrold, 2010). The anterior attention network (AAN) in contrast emerges later than the PAN and is unique to the human brain (Tomasello & Carpenter, 2005). The AAN includes a pathway from the frontal eye fields to the superior colliculus that emerges between three and four months of age (Mundy & Jarrold, 2010) that allows infants to volitionally control eye movement to relevant stimuli, produce and control visual fixations, and alternate eye gaze between others and objects/events (Mundy, 2003). The AAN allows infants to process information internally about their own visual attention which in turn supports spontaneous, volitional, and goal-directed vision and action on objects and people (Mundy & Jarrold, 2010; Mundy & Vaughan Van Hecke, 2008).

The AAN and left-hemispheric frontal-cortical activity are thought to regulate initiating joint attention (IJA) (Mundy, 2017). IJA emerges with the development of the AAN (Mundy et al., 2009) and is defined as the generation and use of *purely social* outputs (i.e., actions, including eye contact and language) to coordinate the attention of others with an object, entity, or event (Mundy & Newell, 2007). This process is cyclical: infants learn to navigate attention towards others who in turn provide social responses and feedback about the surrounding world. Infants seek increasing feedback from others as they gain the ability to attend to multiple sources of stimuli and their subsequent visual attention becomes contingent on this social feedback.

IJA and RJA follow two distinct biobehavioral paths of development beginning at three to four months of age but ultimately converge between nine to 18 months of age to form a distributed and integrated neural network that parallelly processes and integrates self-perception with external social information (Mundy & Jarrold, 2010). In other words, information from the AAN about self-attention, information from the PAN about the attention of others, and information synthesized by distributed neural networks are iteratively processed with greater speed, efficiency, and intricacy, eventually forming a “social executive function” (Mundy et al., 2009; Mundy & Newell, 2007). This system, and capacity to generate internal representations, eventually shapes early social cognitive development including the capacity to adopt multiple perspectives (Tomasello & Carpenter, 2005) and the development of symbolic thought and rule-based attention and reasoning (Frye et al., 1995). In brief, JA is a powerful “social executive function” that emerges in the first months and years of life that supports the development of human connection.

1.3 Joint Attention in Neurodiverse Children and Autism

Many neurodiverse children and children with neurogenetic conditions show difficulties with JA. JA and other forms of communication have shown to be implicated in ADHD (Baird et al., 2000) and children with Down Syndrome show particular difficulty with social referencing (Kasari et al., 1995). Difficulties with JA are most pronounced in autism. Difficulties with JA are a core feature of autism and best distinguish autism from other neurotypes (Charman, 2003). Neurogenetic conditions such as Fragile X Syndrome (Bailey et al., 2008) and Prader-Willi Syndrome (Dykens et al., 2011) have elevated prevalences of autism and children with these conditions therein often exhibit JA challenges. It is therefore vital to assess JA across development because challenges with JA can have cascading effects on development and mental health. Individuals with marked difficulties with JA including autistic individuals tend to have challenges with language acquisition (Carpenter & Tomasello, 2000; Dunham et al., 1993) and social interaction and competence (Mundy & Sigman, 2015), skills necessary for building relationships that can consequently affect quality of life and mental health across the lifespan (Cage et al., 2018; Cooper et al., 2017; Schiltz et al., 2021). Countless pediatric interventions leverage JA as a core skill to target (Kasari et al., 2006; Wright et al., 2013) and JA assessment early on can consequently ensure more optimal outcomes for child wellbeing.

Autism is a neurotype characterized by differences in social interaction abilities and the propensity to engage intensely in focused interests (American Psychiatric Association, 2013). Unlike non-autistic infants, autistic infants do not show as strong a preference for social stimuli (Falck-Ytter et al., 2013; Klin et al., 2009) and instead show visual preferences for non-social stimuli, particularly *objects* in motion, as well as atypical and/or sensorial object exploration (Mottron et al., 2007; Ozonoff et al., 2008). Autistic children show persistent difficulties with JA as a cause or consequence of divergent visual preferences. These difficulties include diminished social orienting and social attention compared to non-autistic children (Dawson et al., 2004) and challenges understanding social signals and their value (Bedford et al., 2012; Parsons et al., 2019). Indeed, by the second year of life autistic children tend to exhibit a lack of, limited, or delayed JA in addition to other forms of social communication such as interactional competencies, symbolic thought, and pretend play (Nyström et al., 2019). Therefore, assessment of JA abilities feature heavily in autism diagnostics (Lord et al., 2000).

The PDPM proposes that social information is not encoded as deeply in autistic individuals (Mundy et al., 2009). The continuous processing and integration of "self- and other-attention" into internal representations seen in non-autistic development does not necessarily occur in autistic development. Instead, autistic children show visual preferences for non-social stimuli (Mottron et al., 2007; Ozonoff et al., 2008). In addition, infants later diagnosed with autism do not seem to gain neurotypical visual attention control (Mundy et al., 2009) and exhibit delays in visual developmental stages (Mundy et al., 2009). The PDPM proposes that biases for non-social stimuli and delays in visual development alter infant active vision to the social feedback of others and explains individual differences in JA challenges in autism, particularly in disengaging attention and integrating facets involved in processing social attention (e.g., social perception, referential understanding) (Mundy et al., 2009). Altered encoding and processing in early life and, in turn, divergent social interactions across contexts have cascading effects on social learning (Bakeman & Adamson, 1984). This is believed to produce a global JA challenge in autism.

Divergent social learning fundamentally alters the presentation of JA in highly specific ways in autism (Baranek, 1999; Goldberg et al., 2005; Jones & Carr, 2004; Mundy & Crowson, 1997; Sigman & Ruskin, 1999). IJA challenges seem to be highly specific in autism such that, compared to non-autistic children, autistic children tend to show markedly less spontaneously initiated bids to coordinate the attention of other persons to objects, entities, or events using social communication behaviors (e.g., eye contact, showing, pointing) (Lord et al., 2000; Morales et al., 2000; Mundy, 2013; Sigman & Ruskin, 1999). Although RJA difficulties may not be specific to autism (Nation & Penny, 2008) (RJA difficulties can be seen in other groups such as ADHD individuals, see Baird et al. (2000)), and limitations with IJA have shown to better differentiate autistic characteristics than RJA difficulties (Charman, 2004; Dawson et al., 2004; Lord et al., 2000), RJA can still be a useful marker of autism in early childhood (Bedford et al., 2012). Rather than difficulties with gaze following and social perception, autism seems to be defined by challenges with "true" RJA, or the ability to engage in referential understanding (i.e., assigning meaning to entities referenced by others) (Bedford et al., 2012; Congiu et al., 2016). Both IJA and "true" RJA require referential understanding. Even if children develop the ability to process "self- and other-attention," this information may not be internalized together to form internal representations that are necessary for referential understanding. Therefore, JA difficulties in autism seems to be characterized by limitations in spontaneously initiated social behavior (IJA) together with challenges with referential understanding (RJA).

Evidence suggests that improved JA can be pivotal in differentiating autistic children and improving outcomes in autism (Charman, 2003) and this is supported by the PDPM. JA difficulties are one of the earliest behavioral features that reliably differentiates autistic children from non-autistic children (Jones & Carr, 2004) and are known to continue throughout childhood and adolescence (Mundy et al., 2017) into adulthood (Pelphrey et al., 2005). Level of JA challenges in autism has shown to be predictive of responsiveness to intervention, “severity” of diagnostic features, and later language and social cognitive outcomes (Dawson et al., 2002; Kasari et al., 2006; Mundy et al., 2016; Watt et al., 2006; Wetherby et al., 2007). Moreover, in autistic children, JA itself can be improved with early and targeted interventions (Kasari et al., 2006; Schertz & Odom, 2007) and improvements are known to lead to positive changes in other domains of development key to social learning (Jones & Carr, 2004), particularly concurrent and future language ability and social interaction and communication (Charman, 2003; Toth et al., 2006). Therefore, reliable JA assessment not only has the potential to enable earlier *and* lifelong detection of autism (compared to tools that measure other autistic features) but could serve as measures of responsiveness to autism intervention.

1.4 Assessing Joint Attention

JA has historically been measured using behavioral assessments that are video recorded for later behavioral coding. For example, the Early Social Communication Scales (ESCS; Mundy et al., 2003) is a common semi-structured behavioral measure of JA that uses stimulating toys and situations to elicit child initiating (IJA) and responding to joint attention (RJA) through eye contact, gestures (pointing, showing, giving, reaching), and language. However, behavioral assessments are limited in that they rely on human coding which can be subjective and time consuming. In contrast to behavioral assessment are *spectral* methods of assessment. Spectral assessment methods are highly acute measurement tools that can provide “spatial or temporal characteristics of participant responses not detectable through observation alone” and are “commonly hailed as the most objective and sensitive metrics for monitoring acute changes,” (Kelleher et al., 2020, p. 2). Eye tracking is one of the most acute spectral methods available to measure human attention and has therefore been a preferred method for studies of JA in the last few decades given the limitations posed by behavioral measures.

In the past half-century, eye tracking has allowed researchers to characterize the development of JA in neurotypical children and how JA may present divergently in other populations, most commonly in autistic children. Eye tracking research has shown that JA processes appear in non-autistic children in the first days and months of life with the development of social perception. Non-autistic infants automatically and robustly orient to social stimuli when they are just days old (Bardi et al., 2011; Simion et al., 2008) that becomes specific to the human face by one month of age (Sanefuji et al., 2014), and show greater pupillary dilation, a measure of information processing, attentional engagement, and emotional arousal (Bradley et al., 2008; Hess & Polt, 1964), when looking at faces (Anderson et al., 2006) compared to other parts of the body (Fitzgerald, 1968). Non-autistic children show difficulty disengaging attention from the face, reflecting deeper processing (Chawarska et al., 2010), and facial recognition has shown to be specific to human faces (Chawarska & Volkmar, 2007), which suggests that infants use the human face uniquely for social learning.

Eye tracking studies have shown that facial viewing in non-autistic children quickly becomes focused on the eyes (Haith et al., 1977; Klin et al., 2002; Maurer & Salapatek, 1976).

Preference for the eyes continues into childhood and throughout the lifespan (Klin et al., 2009). Even two- to five-day-old newborns prefer faces when they maintain direct rather than averted eye contact and demonstrate enhanced neural processing when looking at direct gaze (Farroni et al., 2002). Infants can better discriminate and prefer to look at the eyes by two months of age (Maurer & Salapatek, 1976) and by three months, with the formation of the PAN, infants begin to develop gaze following (Hood et al., 1998) and become better and more efficient at processing gaze direction with experience (Farroni et al., 2002; Mundy, 2013). Longitudinal eye tracking studies have demonstrated that preference for the eyes grows stronger with age. Non-autistic infants look more to human eyes than other core features of the face with this effect increasing from two to six months and remaining stable until 24 months (Jones et al., 2008). As such, non-autistic social perception centers largely on the eyes and between the eyes and other features of the face (i.e., between the eyes and the nose or mouth) (Pelphrey et al., 2002).

Due to altered coding and processing of social information early in life, eye tracking studies have shown that autistic children demonstrate consistent diminished social orienting and engagement that begins in infancy and leads to difficulties with JA in early childhood. Autistic children exhibit diminished attention to and processing of the face and mouth (Chawarska & Shic, 2009) and the eyes (Jones & Klin, 2013), lack of arousal modulation (Chawarska & Shic, 2009) and pupillary dilation (Anderson et al., 2006) when viewing faces, and difficulties with facial recognition (Bradshaw et al., 2011). Social orienting in autism is also modulated by objects (Sasson et al., 2008, 2011) and dynamic versus static video paradigms (Shi et al., 2015) more so than in non-autistic infants. A seminal eye tracking paper found that autistic children do not exhibit preferences for social stimuli (Klin et al., 2009) and this finding has shown to be robust. In a meta-analysis of 38 eye tracking studies comparing the social engagement of autistic and non-autistic children, Chita-Tegmark (2016) found that autistic children tend to exhibit significantly less attention towards social stimuli with a mean effect size of 0.55 across studies. Most recently, eye tracking social engagement in autism has shown to have powerful clinical utility. Two studies published in September of 2023 tested a clinic-based eye tracking measure of social engagement designed to supplement clinical diagnosis of autism in children aged 16-30 months, the EarliPoint Evaluation, and found that the program could designate children with an autism diagnosis as young as 16 months with high sensitivity (71.0%) and specificity (80.7%) (Jones et al., 2023a, 2023b). The Earlipoint Evaluation system is not designed to replace clinical diagnoses but to serve as a measurement tool to aid in a more objective, efficient, and scalable diagnosis.

Eye tracking studies directly testing JA facets, IJA and RJA, in non-autistic infants have primarily focused on mechanisms of social learning (i.e., gaze to the eyes, mentalizing) during JA interactions. Studies agree that non-autistic children seem to develop "true RJA" – gaze following the *eyes* of another while mentalizing– by ten months of age. At ten months, non-autistic children seem capable of understanding another's referential intent (Brooks & Meltzoff, 2005) and exhibit RJA irrespective of whether another's eye gaze is paired with a head turn, indicating the significance of the eyes regardless of head movement during RJA (Thorup et al., 2016). Studies focusing on IJA in non-autistic infants have shown that mentalizing is fundamental not only for social learning but for the development of attentional processes *necessary for* social learning. Specifically, when one-year-olds engage their caregivers in IJA, specifically coordinated triadic gaze, when looking at novel objects, they are significantly more likely to sustain attention *to* that object, reflecting deeper processing (Yu & Smith, 2016).

Eye tracking studies directly testing JA in autistic children have shown normative social perception but diminished overall attention, challenges with referential understanding necessary for "true RJA," and robust difficulties with IJA in autism. When viewing an examiner who looked to a referent object to one side with an equally salient "distractor" object located to her other side, 15-month-olds later diagnosed with autism showed no differences in gaze following to the referent object compared to non-autistic controls but did show less looking time towards the referent *and* distractor objects, indicating decreased attentional engagement overall during RJA situations (Parsons et al., 2019). Another experimental eye tracking study found challenges with referential understanding during gaze following in autistic five-year-olds (Congiu et al., 2016). When watching an actor hide an object in one of two cups across a referential perception condition (cups shuffled, object hidden, and examiner gazed to the correct cup) and a referential understanding condition (examiner gazed to the correct cup only), autistic children performed less accurately than non-autistic controls when forced to rely on their referential understanding of the examiner's gaze without directional cues.

Eye tracking studies showing robust IJA difficulties in autistic children have done so by examining alternate gaze. Alternate gaze is an IJA behavior in which children switch their gaze from an interactive partner to an object, entity, or event with the goal of sharing the experience with that person for social learning or enjoyment (Mundy et al., 2003). Figure 1 displays images commonly used to visualize IJA and alternate gaze by Mundy and colleagues (2003). Thorup and colleagues (2018) used eye tracking to examine the developmental trajectories of alternate gaze in 51 infants at elevated likelihood for an autism diagnosis (ELA) and 16 non-autistic infants at 10 and 18 months and found that ELA 10-month-olds demonstrated atypical alternate gaze when interacting with adults. In addition, less alternate gaze at 10 months was correlated with more autistic characteristics and less JA behaviors (specifically showing and pointing) on the ADOS-2 at 18 months of age.

Similarly, Nyström and colleagues (2019) examined IJA in 81 ELA and 31 non-autistic infants followed at 10-, 14-, and 18-months by testing group differences in alternate gaze frequency and trajectory on a novel experimental IJA task. Infants viewed an examiner who gazed back at them, with a large light each to her right and left. While maintaining infants' gaze, one of the lights would flash and infant alternate gazes between the examiner and light signaled that the infant was "checking in" to share or to see if the examiner was *also* seeing the strange flashing light. Nyström and colleagues (2019) found that infants later diagnosed with autism had lower rates of alternate gaze and positive alternate gaze slopes across time compared to non-autistic infants and ELA infants who were not diagnosed with autism, who had higher rates of alternate gaze and negative alternate gaze slopes (likely corresponding with the development of more advanced IJA behaviors with age, resulting in less alternate gaze use). In summary, eye tracking has been used for several decades in pediatric populations as an acute measurement device to show pertinent difficulties with JA in autistic children and in ELA infants.

1.5 Remote Neurodevelopmental Assessment Tools

The initiation of the COVID-19 pandemic in 2020 introduced unparalleled barriers for *any* child seeking neurodevelopmental assessment. Global shutdowns barred in-person procedures and prohibited nearly all existing assessments for almost two years. Luckily, with advances in telemedicine and technology over the last two decades, together with an urgent need to serve families seeking diagnoses remotely during COVID-19, the field of remote neurodevelopmental assessment has quickly burgeoned.

Remote telehealth assessment tools are a critical next step in pediatric assessment. Racially and ethnically minoritized individuals with and without disabilities are underrepresented in research and are more likely to experience barriers to research participation such as rurality (George et al., 2014). However, studies show that minority families are just as willing as Non-Hispanic white families to participate in research (Wendler et al., 2006). Remote tools could diversify research populations immensely by bringing in families willing to participate who are otherwise unable to travel to clinics due to location, cost, availability of childcare, or time constraints. Remote research is also a powerful tool in reaching large numbers of certain clinical populations. Telehealth research and services are often the only way to reach families with children with rare neurogenetic syndromes who are geographically diverse and often have co-occurring conditions that make travel challenging (Hyde et al., 2020; Kelleher et al., 2020). Families also seem to like telehealth. One study of a remote intensive intervention for autistic children found that caregivers had high participation rates and satisfaction with the method and low-income caregivers tended to have the most satisfaction (Wood de Wilde et al., 2023).

Finally, bringing quality assessment tools home could substantially increase assessment data quality. Clinic and lab-based assessments are not inherently naturalistic. Studies often do not acquire data from children otherwise willing to participate due to healthcare-setting-induced anxiety and a lack of control of their environment (Lerwick, 2016). There are many challenges to in-person developmental research as well. Children may have difficulty cooperating behaviorally, may tire or become upset more easily, and may not feel comfortable being led through study procedures by unfamiliar persons. In-person and particularly behavioral assessments are also subject to bias. Clinician implicit racial and ethnic biases are strongly related to health outcomes (Hall et al., 2015). Non-autistic (Fadus et al., 2020) and autistic (Mandell et al., 2007) racial and ethnic minority children are more likely than Non-Hispanic white children to receive classifications of disruptive behavior disorders during diagnostic evaluations. Bringing quality assessment tools home would give families the opportunity to have their children participate in research in a comfortable, familiar setting. Without the time and space constraints of being in a clinic or lab, children can be guided by their primary caregivers, go through study procedures at their own pace, and have the flexibility to try again later. Hereafter, we give a brief overview of existing remote neurodevelopmental assessment tools, their benefits, and limitations.

The earliest forms of remote assessment involved coding old home videos of young children retrospectively, usually filmed during typical milestones (i.e., birthday parties), for behavioral autistic features related to sensory processing, general social behaviors, communicative gestures, behavioral regulation, and JA (Baranek, 1999; Osterling & Dawson, 1994; Watson et al., 2013). Although these methods were able to successfully discriminate between children with and without an autism diagnosis with high classification rates, coding post hoc can be less than ideal. Tasks are not standardized by any means and families without videos of their child in early life (or video recording equipment) are not studied. Retrospective coding can also be exceptionally time consuming in that it requires intensive training to achieve inter-rater reliability (IRR) and behavioral coding is still susceptible to human error even when IRR is achieved. Retrospective coding was therefore a first step in remote neurodevelopmental assessment.

Several newer remote neurodevelopmental assessments are designed to be filmed and led by parents, who are often coached live by trained clinicians or researchers, to elicit features of autism at home and to be scored live or post-visit by trained clinicians (Dow et al., 2020, 2021; Nazneen et al., 2015; Smith et al., 2017). Most tools use a combination of live and *store-and-forward methods.* Store-and-forward remote methods are those in which families film videos of the participant live to later be scored by clinicians or research groups (Alfuraydan et al., 2020).

The most common remote neurodevelopmental assessments are the Systematic Observation of Red Flags (SORF; Dow et al., 2017, 2020), the Naturalistic Observation Diagnostic Assessment (NODA; Nazneen et al., 2015; Smith et al., 2017), the Brief Observation of Symptoms of Autism (BOSA; Dow et al., 2021), and the TELE-ASD-PEDS (TAP; Corona et al., 2020, 2021).

The SORF is a common store-and-forward neurodevelopmental assessment that generates a six-item composite score shown to be reliable, sensitive, and specific in classifying 16- to 24-month-olds as autistic, non-autistic, or developmentally delayed (Dow et al., 2017, 2020). SORF caregivers interact with and film their child during regularly occurring, everyday activities (e.g., snack time, playtime) for one hour. Videos are later coded by undergraduate research assistants (RAs) on a four-point scale for 22 red flags associated with DSM-5 diagnostic criteria, including the six composite score items: (1) measuring eye gaze to faces, (2) showing and pointing, (3) coordination of nonverbal communication, (4) interest in people compared to objects, (5) repetitive object use, and (6) excessive interest in particular objects, actions, or activities. The NODA is another store-and-forward tool, a mobile app, that is a reliable and clinically useful assessment for children 18 months to just under seven years old (Nazneen et al., 2015; Smith et al., 2017). The NODA app leads caregivers through videorecording and administration of four 10-minute activities (a mealtime, play time with others, play time alone, and general caregiver concerns) while their mobile device is mounted on a tripod. Videos are subsequently uploaded and scored (together with scores from caregiver-reported developmental histories) using DSM-5 criterion by raters with at least ten years of experience conducting autism assessments. However, the NODA is not yet specific in differentiating autistic children from non-autistic children with developmental delays (Dow et al., 2021).

The BOSA is another promising 12- to 14-minute remote neurodevelopmental assessment that is adapted from ADOS-2 modules and the Brief Observation of Social Communication Change (Grzadzinski et al., 2016) that has shown to be sensitive in discriminating autistic individuals from non-autistic individuals ranging from 15 months to 42 years, regardless of language and cognitive ability (Dow et al., 2021). For child diagnoses, BOSA modules utilize ADOS-2 materials for a sequence of filmed activities based on the child's language level and age (i.e., blowing bubbles, free play, Slap Jack) in which the caregiver is instructed by an examiner to engage and play with their child. The interactions can be scored live or with store-and-forward videos by ADOS-2- and BOSA-reliable clinicians or researchers using ADOS-2 scoring and the DSM-5 checklist. Finally, the TAP, designed using machine learning techniques, is a 15- to 20-minute store-and-forward assessment for children under 36 months that uses clinician-led caregiver coaching on a HIPAA-compliant video platform in the administration of eight activities and social bids (Corona et al., 2020, 2021). ADOS-2-trained clinicians give a diagnostic impression (e.g., autism or no autism) following the assessment and rate seven behaviors using dichotomous and Likert scales, which are pooled to create a sum score. Although the TAP has demonstrated parental and clinician feasibility, the TAP currently has no psychometric properties available.

Although useful, the SORF, NODA, BOSA, TAP, and other existing remote neurodevelopmental assessments have several notable limitations. Many assessments are time consuming as it takes at least one hour to complete the SORF and at least 40 minutes to complete the NODA which can be burdensome for families with small children. The TAP and BOSA are advantageous here in that they are much shorter (between 12-20 minutes) although the BOSA requires a large set of ADOS-2-specific materials (which are expensive and can only be purchased by clinicians). Another limitation is the need for other child information that may not necessarily be obtainable through telehealth (i.e., measures of cognitive development) that may influence the scoring of tools like the BOSA. Filming can be a further limitation. The SORF has only used videographers to film sessions which is not a naturalistic nor feasible way to conduct research with families, particularly if videographers need to service families during pandemics or living in rural areas. The NODA has caregivers film interactions on a smart phone app, while the BOSA and TAP film videos using telehealth platforms. Telehealth and app approaches are advantageous in that they can be HIPAA-compliant, accessed for free via smart phone, computer, or tablet, and can allow for mobile filming. However, these approaches can be limited in that videos may vary in viewing and sound quality.

Most importantly are the following three major limitations. First, two of the four tools have yet to be validated *for use remotely*, the purpose of such assessments. The TAP has only been tested preliminarily as a remote assessment and has yet to be validated in the home setting. Visits have only occurred in research laboratories and have not occurred without live, in-person, caregiver coaching. While Dow and colleagues (2021) examined 307 participants with 453 BOSA assessments, only 20 BOSAs were actually done remotely with telehealth and the remaining were completed by caregivers or clinicians in a laboratory setting. Second, these tools often require advanced training for scoring. The NODA requires scorers to have experience conducting autism assessments, while the BOSA and TAP necessitate ADOS-2-trained clinicians for scoring. ADOS-2 training is expensive, requires attendance at a multi-day workshop (typically in person), is limited to experienced autism professionals and clinicians, and obtaining clinical or research reliability can be noticeably prolonged, requiring time-intensive and supervised training (Charman & Gotham, 2013; Zander et al., 2016). Only the SORF has a scoring system (e.g., behavioral coding done by RAs) that is not reliant on years of training.

Third and finally, these tools are almost exclusively observational in nature, in direct contrast to many in-person assessments and spectral methods like eye tracking, heart rate variability, electroencephalography, etc., that can more accurately parametrize autistic features. Exclusively observational assessments can be inherently subjective and even biased as they are reliant on behavioral measurement and coding as well as video viewing and sound quality. These observational paradigms also depend on caregivers to choose and/or administrate assessment activities. Although more naturalistic, caregivers of children who struggle to engage in joint social interaction may employ extra strategies or scaffolding behaviors to coordinate their child's attention (Chawarska et al., 2013), which may not give clinicians a true picture of the child's strengths or difficulties with JA.

A select few studies have applied automated coding methods to parametrize JA during remote observational assessment more acutely and objectively. One research group at Duke University has piloted a series of screen-based apps in which children are videorecorded while watching simple social scenes and videos and are later coded using automated computer vision coding (Egger et al., 2018; Hashemi et al., 2016) to quantify social behavior and gaze. Hashemi and colleagues' scalable app was designed for use in home, primary care, and school settings. The app involves videorecording children while they watch recorded scenes (i.e., a mechanical bunny, bubbles) while seated in front of their caregiver, during which an examiner attempts to engage the child. The authors employed frame-by-frame computer vision, a machine-learning algorithm, to code videos for social referencing (e.g., head turns toward the parent or examiner) and facial expressions (difficulties with which can be indicators of autism) in twenty 16- to 30-

month autistic and non-autistic children to quantify social communication behaviors. The app was shown to be feasible and as accurate as human behavioral coding in quantifying social referencing and positive affect. However, the app has only been piloted in twenty children and has yet to be tested outside of a clinical setting.

The same research group has since developed the groundbreaking Autism and Beyond App (Egger et al., 2018) that uses the same computer vision software to measure facial expressions and attention. The app has been extensively tested in 1756 families with children aged 12 to 72 months. Caregivers download the iPhone- and iPad-compliant app and go through an automated consent. If eligible, they complete surveys and their child watches four short movies on the caregiver's iPhone or iPad while the device videorecords the child's face. The entire experiment is conducted via app without person intervention, has proven to be a feasible means to collect meaningful JA data, and has shown significant differences in facial expressions and attentional engagement according to autism status. Autistic children exhibited significantly less range in facial and attentional engagement than did non-autistic children.

Although automated coding and computer vision are more *acute forms* of behavioral measurement, these apps still utilize *behavioral observations and measurement*. To our knowledge, only one study has gone beyond the scope of solely behavioral measurement to employ multimodal, spectral tools for remote assessment. The Parent-Administered Neurodevelopmental Assessment (PANDABox; Kelleher et al., 2020) uses video-recorded observations and survey data coupled with the LENA recording system (Xu et al., 2008) to record child sounds and vocalizations and a portable Actiwave ECG monitor to examine child psychophysiology. This multimodal spectral approach allows PANDABox researchers to leverage multifaceted data analysis, including minute-by minute voice recording and a biomarker for child physiological arousal, to make the best-possible diagnoses. Nevertheless, no remote neurodevelopmental assessments exist that combine observational paradigms together with eye tracking. This is an essential next step as eye tracking is now one of the most precise and reliable methods with which to capture attentional metrics in even the youngest of children. Difficulties with JA, an *attentional* process, are a core feature of autism, and in-person assessment tools often employ eye tracking. Including eye tracking in remote assessment is a critical next step to more accurately quantify JA difficulties in order to ultimately expedite the pathway for a reliable autism diagnosis.

1.6 The Potential for Portable, Remote Eye Tracking

Remote neurodevelopmental assessments have not yet leveraged the powerful technology of eye tracking. Furthermore, very few research studies exist that have utilized eye tracking remotely because few portable eye trackers exist. Nearly all eye tracking studies to date use stationary, lab-based eye trackers that can be cost prohibitive for research teams and require a high degree of set up (space, additional computing needs). The few portable eye trackers that are available have been created largely to serve as augmentative and alternative communication devices (AACs; see the Tobii Dynavox products here) for individuals with disabilities. As such, these tools are not meant to produce quantifiable data for research purposes. However, the few studies that have used portable eye tracking to measure JA in the field have been promising.

To our knowledge, only one study has directly tested JA in children remotely using portable eye tracking. Benson-Goldberg & Erickson (2021) examined the possibility of using a portable Tobii PCEye Go eye tracker (an AAC device) to assess responsiveness to an emergent literacy reading intervention in a three-year-old with CDKL5 at home. The authors used eye tracking during a shared book reading task to test how well the child engaged in RJA when her mother pointed to book pictures. The study demonstrated that remote and portable eye tracking was a reasonable method for acquiring JA data collected at-home and by caregivers. Guided by remote support from a researcher, the child's mother was able to successfully set-up, calibrate, and use the eye tracking technology. The portable eye tracker was also able to provide simple outcome data on the efficacy of the intervention measured with Tobii Dynavox Gaze Viewer application software (*Gaze Viewer*, 2022).

Benson-Goldberg and Erickson's (2021) study was the first to provide proof of concept that remote and portable eye tracking can be done successfully by families at home. While informative, the study was hindered by the lack of automatically produced data. Gaze Viewer, one of only several systems that can analyze portable eye tracking data, only provides fixation counts and video recordings of fixation points which must be manually coded. A silver lining of the COVID-19 pandemic was the fact that researchers and eye tracking companies alike recognized the need for portable, *research-grade* eye trackers for research outside of the lab and many have since been developed (see here and here), particularly eye trackers that are compatible with mobile electronics like laptops. Researchers have since capitalized on these devices. Although the research was completed physically within a laboratory, one study brought eye tracking technology to the remote country of Bhutan with the use of a portable Tobii Pro Nano device to examine the relationship between maternal postpartum depression and infant JA development (Astor et al., 2022). Remote, portable eye tracking is the clear next step for remote JA assessments as it has the potential to bring research-grade measures of JA home to families awaiting future autism diagnoses, ultimately expediting the pathway to diagnosis.

Another limitation of Benson-Goldberg and Erickson's (2021) study was the lack of quantifiable measures that would have specifically answered the question: "is completing portable eye tracking of JA at home through remote means truly feasible?" Feasibility studies can solve this limitation. Feasibility studies are a standardized way in which to assess whether a novel research method can be reasonably completed within specific populations or settings (Bowen et al., 2009). In other words, they answer, "Can it work?" Bowen and colleagues (2009) specify eight variables relevant to testing new research methods or research methods that have been piloted and require additional testing: acceptability, adaptation, demand, expansion, implementation, integration, limited-efficacy testing, and practicality. Research is needed to answer the question: "Does remote and portable eye tracking of JA actually work?" Portable and research-grade eye tracking technology, combined with *quantifiable* measures of feasibility, would provide answers to this question.

1.7 The Current Study

The current study tests the feasibility, utility, and validity of the TRAVEL Study, a remote and portable eye tracking and store-and-forward behavioral study designed to measure early JA in young children aged 30 months to 54 months. Most screening and assessment tools are tailored for two- and three-year-olds (autism can now be diagnosed as young as sixteen months (Jones et al., 2023a, 2023b)). Thirty months is the oldest age at which the American Academy of Pediatrics recommends developmental assessment at well-child visits (Richerson et al., 2017). However, tools are needed to identify autistic features in children who may face barriers to the earliest screening opportunities and who may "fall through the cracks" without further developmental assessment opportunities. Therefore, we piloted this program for the purposes of ultimately capturing preschool children who may be missed during the diagnostic process and with the transition to school at about five years of age, we studied children from 30 months up until 54 months. Importantly, children within this study already had autism diagnoses so as to determine typical profiles of JA in autistic and non-autistic children. The current study has three aims.

First, this study aims to assess the feasibility of TRAVEL as a remote, portable eye tracking and behavioral JA assessment by investigating three feasibility indicators: demand, acceptability, and practicality. Second, we aim to determine whether TRAVEL can accurately discriminate between JA behaviors in autistic and non-autistic children using two outcome measures created with the following analytical approaches. The first approach uses logistic regression to examine whether child diagnosis predicts the functional form of JA – *TRAVEL trajectories* – during eye tracking in response to specific "events," one designed to elicit RJA and, the other, IJA. The second approach employs structural equation modeling (SEM), specifically exploratory factor analysis (EFA), to estimate children's global JA profiles - *total TRAVEL scores* – by pooling JA scores from TRAVEL portable eye tracking and behavioral measures. Third, we aim to test the construct validity of TRAVEL outcomes. Convergent validity of TRAVEL outcomes will be assessed by examining their associations with (1) validated caregiver-reported measures of JA and (2) developmental measures commonly associated with JA: expressive and receptive language. Divergent validity of TRAVEL outcomes will be examined by analyzing their associations with caregiver-reported measures of the restricted and repetitive behaviors and interests factor score on the SRS-2 (SRS-RRBs). Restricted and repetitive behaviors and interests have previously been used to test the divergent validity of JA measures (Nowell et al., 2018) because they have shown to be largely theoretically independent of JA (Mundy et al., 1994). Caregiver-reported measures of cognitive development will also be measured and used as a covariate in study analyses.

If the TRAVEL Study proves to be a feasible, useful, and valid way in which to measure JA in children at home, we ultimately hope to develop TRAVEL as a remote assessment tool and outcome measure to reach broader populations for pediatric research. In addition, TRAVEL would be a proof of concept for remote, portable eye tracking research on JA in neurodiverse pediatric populations that could broaden the scope of eye tracking research altogether and substantially diversify research populations. Our research aims and hypotheses are as follows:

Aim 1: Test the feasibility of the TRAVEL Study for caregivers by examining three indicators of feasibility: demand, acceptability, and practicality.

Hypothesis 1:

H1a. <u>Demand</u> was defined as the likelihood of "program use" or the likelihood that individuals would participate in the TRAVEL Study. We hypothesize that TRAVEL will demonstrate sufficient demand to warrant further study.

H1b. <u>Acceptability</u> was defined as the extent to which the TRAVEL Study was suitable for the intended participants. We hypothesize that caregivers will deem the TRAVEL Study and its remote delivery and portable eye tracking procedures acceptable.

H1c. <u>Practicality</u> was defined as the extent to which the TRAVEL Study could be reasonably carried out with the intended participants using existing resources, means, and circumstances. We hypothesize that TRAVEL will prove a practical way in which to acquire behavioral and portable eye tracking data on JA remotely in autistic and non-autistic children.

Aim 2: Test whether TRAVEL outcome measures (TRAVEL trajectories and total TRAVEL scores) differentiate autistic and non-autistic children.

Hypothesis 2: We hypothesize that both outcome measures will differentiate autistic and non-autistic groups such that autistic children will (1) exhibit significantly different *TRAVEL trajectories* than non-autistic children in that their functional forms of RJA will have a significantly smaller slope and their functional forms of IJA will take the shape of a polynomial as opposed to a linear spline (H2a); and that autistic children will have (2) significantly lower *total TRAVEL scores* than non-autistic children (H2b).

Aim 3: Analyze the construct validity of TRAVEL outcome variables (total TRAVEL scores and TRAVEL trajectories) by investigating whether they are associated with and predictive of validated caregiver-report measures of (a) JA, (b) measures of child development commonly associated with JA (expressive and receptive language), and (c) RRBs.

Hypothesis 3: We hypothesize that TRAVEL outcomes will be moderately to strongly associated with and significantly predictive of validated caregiver-report measures of JA (H3a) and of expressive and receptive language (H3b). We hypothesize that TRAVEL outcomes will be weakly associated with and not significantly predictive of RRBs (H3c).

CHAPTER 2: METHODS

All study procedures were approved by the UNC Chapel Hill Institutional Review Board.

2.1. Sample

Participants in the TRAVEL Study were 50 children and their primary caregivers living in North Carolina. The sample was made up of 25 children with a caregiver-reported diagnosis of autism and 25 non-autistic controls. We opted to study children aged 30 to 54 months to test the feasibility of remote eye tracking in young children with an autism diagnosis, with the ultimate goal of extending this research to *younger* children *seeking* an autism diagnosis in the future. Participants were recruited via the Carolina Institute for Developmental Disabilities Autism Research Registry and Child Development Research Registry, social media posts in autism-specific parent groups on Facebook (e.g., Autism Society of North Carolina chapter groups), and with paper flyers. Caregivers completed a phone assessment prior to the visit and were deemed eligible if they met the following criteria: (1) resided in North Carolina; (2) had a child between 30 to 54 months of age; (3) were willing to complete a remote telehealth visit using supplies mailed or hand-delivered; (4) were willing to have their child video-recorded using Zoom (HIPAA-compliant); and (5) had WIFI in their home. Caregivers completed a verbal consent on the phone assessment.

Children were 85.7% white, 4.1% Black/African American, 2.0% Asian, 6.0% Multiracial, and 2.0% (of caregivers) preferred not to answer, and 10.2% were Hispanic/Latinx. Children were on average 47.0 months of age (SD = 6.86). Non-autistic children were 48% female. Autistic children were 32% female (see Table 1). Caregivers were 83.7% white, 6.1%

Black/African American, 4.1% Asian, 2.0% Multiracial, 2.0% “Other” race, and 2.0% preferred not to answer, and 8.2% were Hispanic/Latinx. Caregivers were 93.9% female, and the majority (95.8%) had a bachelor’s degree or higher, were employed (83.3%), and had an annual household income of $75,001 or higher (73.0%). Fourteen families received their TRAVEL Box by mail and the remaining (N=36) received theirs by drop off. An independent samples t-test showed that families with an autistic child were significantly more likely to be mailed study supplies compared to families with a non-autistic child, $t(41.38) = -2.62$, $p = 0.01$.

2.2. Procedures

Participating families received their TRAVEL Box with study supplies prior to their visit, which included a ThinkPad laptop computer, a Tobii Pro Nano portable eye tracker, toys for caregiver-child tasks, and a detailed set-up instruction booklet (the TRAVEL Guidebook). Study families located within driving distance from the study lab were hand-delivered their TRAVEL Box in addition to a wooden chair and tray table to standardize child seating position during eye tracking. Study families not within driving distance were mailed their TRAVEL Box and were told to seat their child in a highchair, car seat, or on their lap but were given tools to standardize seating (e.g., a 60cm ruler and visual instructions). Using the TRAVEL Guidebook, caregivers prepared for the study visit by logging into the study computer, connecting to their home WIFI network, placing the screen-mounted portable eye tracker, and connecting to the study Zoom room. From there, their visit leader provided an overview of the study, answered any questions, obtained verbal consent once again, and began Zoom recording.

To begin the study visit, the visit leader established remote control of the participant screen and began to set up Tobii Pro Lab (Tobii Pro AB, 2014) for portable eye tracking. During this time, caregivers were told to see the TRAVEL Guidebook for instructions about seating, measuring their child's distance from the laptop to 60cm using the provided ruler, and telling their child that they were going to watch some special videos. Once all technical systems were in place, the visit leader calibrated the child's eyes using a nine-point calibration method in which they attended to a red dot that moved to nine locations on the computer screen. Caregivers were told to instruct their child to watch the dot on the screen and the visit leader coached the dyad during calibration. Calibration was repeated as needed if failed.

Subsequently, children participated in about nine minutes of portable eye tracking and were video recorded via the Tobii Pro Lab system for later viewing. Following eye tracking, dyads completed two caregiver-led behavioral tasks, a task measuring children's response to name and a parent-child interaction. After visit completion, caregivers were instructed to exit out of Zoom and to return all items to their TRAVEL Box. Zoom videos were stored on the visit leader's computer on a secure server immediately after data collection. At the family's convenience, a study RA returned to pick up the materials or, for non-local families, a mail carrier was scheduled to pick up their study box at their home. Post-visit, caregivers were sent a unique link to online study surveys via UNC Qualtrics (HIPAA-compliant). Caregivers were compensated $40 in cash for their participation.

2.3 Measures

2.3.1 Eye Tracking

Children were eye tracked using the Tobii Pro Nano portable eye tracker at a sampling frequency of 60 Hz while watching five eye tracking paradigms (see Figure 2). The first task was *Story Time*, a three-minute video in which an examiner read a children's book, *The Lorax*, while pointing to images using a time delay (gasp) on every other page. *Story Time* was a warm-up activity meant to engage children and served as one measure of RJA. Proportion scores were calculated by dividing the number of times children “correctly” followed examiner points by the total number of points that occurred during *Story Time* when children were looking anywhere on the screen. Trials in which children were not attending to the screen were not counted and this is the case for all subsequent proportion scores. See Table 2 for descriptive statistics of individual TRAVEL Items by child group.

Children viewed four remaining tasks in which the examiner was seated between two items to her left (L) and right (R) (see Figure 3 for a timeline of portable eye tracking tasks). These tasks were shown and then repeated in the same consecutive order to account for data loss due to movement or inattention, but from R to L for the second set. During *Puppets*, the examiner sat between white screens, each with a hole for a puppet. The examiner carried out a series of three pointing presses to each side, first a time delay and point to the puppet, second, saying, “Look!” while pointing to the puppet, and third, saying, “Look! A sheep/cow!” while pointing to the puppet. Proportion scores were calculated by dividing the number of times children “correctly” followed examiner points (measured as within an area of interest (AOI) created in Tobii Lab Pro prior to data analysis) by the number of total points during each trial. During *Robots* and *Rabbit* (based on the light flashing task in Nyström et al. (2019)), the examiner began by attracting children’s attention (saying hello and waving) before sitting silently while a remote-controlled robot/toy rabbit was activated one at a time on each side. Similarly, during *Balloons*, the examiner attracted children’s attention before blowing up a balloon, holding it to each side one at a time, and then slowly releasing air from the balloon to create a funny sound. These tasks examined IJA by measuring the number of times children alternately gazed between the examiner and the unexpected event (robots spinning, rabbit hopping, balloon deflating) to “check in” with the examiner. Proportion scores were calculated by dividing the number of children's gazes to the examiner by the total amount of time the object is active during each trial.

2.3.2 Behavioral Tasks

Children completed two behavioral tasks videorecorded on Zoom that were led by the caregivers and facilitated by the examiner. The first task measured children's response to name (*RTN*), a form of RJA. Children were seated in front of the computer screen and were given a "distractor" toy (children were allowed to choose between a vibrating squirrel toy, a red fire truck, or both with the goal that children would choose the toy in which they were most interested). Caregivers were instructed to stand to the front left (L), front right (R), behind L, and behind R of their child (see Figure 4) and at each position, to call their child's name twice loudly with a five second pause in between name calls. Caregivers praised children for "correct" responses so they would not begin ignoring name calls. Caregivers remained silent and neutral when children did not respond to name calls. *RTN* proportion scores were calculated in which the number of "correct" responses to name (defined as a head turn *and* eye contact toward the caregiver after each name call) were divided by the total number of name calls. IRR was determined using an intraclass correlation generalizability (*g*) coefficient (ICC; Cicchetti, 1994). RA's coded *RTN* tasks and IRR was established by obtaining an Intraclass Correlation Coefficient (ICC) of α=.80 reliability with the master coder, the author of this study (JG), on five initial training videos.

The second behavioral task was a 10-minute parent-child interaction (*PCX*) designed to measure frequency of discrete eye contact and gestural IJA behaviors. The caregiver-child dyad was instructed to play with a standardized set of toys (Lego Duplo blocks, wooden dolls, wind-up toys, play dishes/food, a puzzle, Minnie and Mickey Mouse stuffed animals, and baby doll accessories) wherever they were most comfortable, either seated at a table or on the floor. Videos were later behaviorally coded for children's spontaneous use of alternate gaze (or two-point looks between the caregivers and toy or vice versa) and gestures (showing, giving, and pointing). An overall ICC of 80% for *PCX* coding was established between JG and undergraduate RA with five training videos. For specific codes, excellent reliability for alternate gaze was achieved with an ICC of α=.90 and good reliability was achieved for gestures with an ICC of α=.70.

2.3.3. Demographic Survey

Caregivers completed a brief demographic survey when beginning study questionnaires. The demographic survey provided information specific to caregivers (e.g., highest level of education, income) and children (e.g., current age, race, and (for autistic children) information about their autism diagnosis).

2.3.4 Caregiver report of autistic characteristics, social communication, and social competence

The Social Communication Questionnaire-Current (SCQ; Rutter et al., 2003) and the Social Responsiveness Scale Preschool Form, 2nd edition (SRS-2; Constantino & Gruber, 2012) were used to measure caregiver-reported JA behaviors in the form of social communication, social competence, and autistic characteristics.

The SRS-2 (Constantino & Gruber, 2012) is a common, 65-item, caregiver report measure that identifies the presence and intensity of autism characteristics, including social difficulties and repetitive behavior patterns, in autism and other neurotypes. The SRS-2 preschool form can be completed in 15-20 minutes and is designed for children aged 30 to 54 months of age. Caregivers report on their child's behavior over the past six months and rate items on a four-point Likert scale from "not true" (1) to "almost always true" (4). The SRS-2 yields five domain scores, Social Awareness, Social Cognition, Social Communication, Social

Motivation, and Restricted, Repetitive Behaviors and Interests, which cluster to form a two-factor structure: Social Communication and Interaction (the first four social items) and Restricted, Repetitive Behaviors and Interests (the last item). Composite *T*-scores yield an overall Total *T*-score and a score of greater than or equal to 60 indicates elevated autism characteristics, with higher scores indicating greater "severity" of autism features. The psychometric properties of the SRS-2 are strong, with standardization samples of autistic children yielding an internal consistency reliability of α=.95, sensitivities of α=.74 to .85, and specificities of α=.69 to 1.00 in autism (Bölte et al., 2011; Moody et al., 2017).

The SCQ-Current (Rutter et al., 2003) is a common 40-item caregiver report assessment tool that evaluates communication and social skills in children who are or who are suspected of being autistic and accounts for child behaviors over the most recent three-month-period. Caregivers respond to items with dichotomous yes or no responses. Question one probes caregivers for child language ability with the question, "*Is she/he now able to talk using short phrases or sentences?"* and is not scored. If caregivers have a speaking child and answer "yes" to question one, they are assigned all 39 remaining items including six questions about their child's language ability. If caregivers have a non-speaking child and answer "no" to question one, they are not assigned the six language ability questions and complete the remaining 33 items. Each item receives a value of 0 for the absence of and 1 for the presence of "abnormal behavior," so speaking children can score between 0 and 39 points whereas non-speaking children can score between 0 to 33 points. A cut-off score of 15 indicates the presence of autism, although newer evidence suggests that a score greater than or equal to 10 may be more accurate (Barnard-Brak et al., 2016). The sensitivity of the SCQ in correctly classifying autistic children is α=.96 and the specificity is α=.80 in samples without intellectual disability, although the specificity is known to drop for children with an intellectual disability (Moody et al., 2017).

2.3.5 Caregiver report of language and cognitive development.

Caregivers rated their children's receptive and expressive language, which are developmental factors often associated with JA, via the MacArthur Bates Communicative Development Inventory (MCDI; Fenson et al., 1993). The MCDI (Fenson et al., 1993) is a widely used caregiver report instrument which captures information about children's developing abilities in early language, including vocabulary comprehension, production, gestures, and grammar, and has been used to describe language profiles in autism (Charman et al., 2003). The MCDI has two parts: Part I, Early Words, and Part II, Actions and Gestures. Part I is divided into four sections: First Signs of Understanding (3 yes/no items), Phrases (28 yes/no items), Imitation and Labeling (2 three-point Likert scale items from "not yet" to "often"), and Vocabulary Checklist (396 items with answer options of "does not understand," "understands but does not say," and "understands and says," organized into 19 semantic categories). Part II is divided into five sections: First Communicative Gestures (12 three-point Likert scale items from "not yet" to "often"), Games and Routines (6 yes/no items), Actions with Objects (17 yes/no items), Pretending to Be a Parent (13 yes/no items), and Imitating Other Adults Actions (15 yes/no items). For each item, caregivers can also choose not to answer. The MCDI produces scores for Expressive Vocabulary and Receptive Vocabulary with a range of possible scores from 0 – 89. The MCDI has an internal consistency of α=.67 to .96 and test-retest reliability of α=.86 to .95 (Dale et al., 1989).

Caregivers also rated their children's cognitive development with the Motor and Social Development questionnaire (MSD; Peterson & Moore, 1987). The MSD is a 48-item caregiver report survey developed by the National Center for Health Statistics for the National Longitudinal Survey of Youth to measure the motor, social, and cognitive development of young children. Items are derived from standard measures of child development (i.e., the Bayley Scales of Infant Development; Bayley, 1993) that have high reliability and validity. Based on the child's chronological age, caregivers answer 15 age-appropriate items out of 48 items dichotomously with “no” (0) or “yes” (1), each representing a developmental milestone that the child has or has not achieved. Total raw scores are obtained by summing the 15 items. Due to the fact that items are successive in the MSD, the instrument does not have a Cronbach’s Alpha Coefficient. However, the MSD has several characteristics suggesting strong psychometric properties, including a very low chance of selective non-response and that large samples of children have yielded overall intelligence quotient (IQ) means close to 100 (the average IQ) (*NLSY Child Handbook: A Guide & Resource for the NLSY 1986 Child Data (PDF)*, 1989). Cognitive development scores on the MSD were used in study analyses as a covariate.

2.3.6 TRAVEL Caregiver Survey

Caregivers were given a 27-item survey after their TRAVEL Study visit that has been updated based on protocol used and graciously shared by Ravindran et al., (2019) (for the full survey, see supplemental materials). The survey assessed caregiver opinions and acceptability of the general study. For example, caregivers were asked questions such as, “*How was your experience with delivery and pick up of the TRAVEL Study materials?*” and answered according to a five-point Likert scale from “easy” (-2) to “difficult” (2). The survey also contained checks for the TRAVEL eye tracking tasks such as, “*Did the participant tolerate the eye tracking session?*”, “*In your opinion, did the participant enjoy their eye tracking session?*”, and “*Did the participant exhibit excessive movement, fidgeting, or changing from sitting to standing?*”.

2.4 Analytical Strategy

2.4.1 Feasibility (Aim 1)

Feasibility of the TRAVEL Study was assessed according to three of eight domains suggested by Bowen et al. (2009) as noted above: demand, acceptability, and practicality.

2.4.1.1 Demand. Demand was assessed as the total number of participating families who were eligible and who enrolled and consented (regardless of whether they completed the study) relative to the total number of families who had knowledge of the TRAVEL Study. This ratio was calculated by dividing the number of enrolled and verbally consented families by the number of families who likely had knowledge of the study through each trackable recruitment source. The Carolina Institute for Developmental Disabilities Autism Research Registry and Child Development Research Registry provided our study team with the total number of families emailed and who therefore had knowledge about our study. Metrics were also provided by Facebook, which tracks the number of individuals who saw the post. The total number tallied across all platforms was used to create this ratio and these numbers were used to sum the number of families likely knowledgeable about the study on Facebook.

2.4.1.2 Acceptability. Acceptability was measured as total score from 0 to 12 when answering three TRAVEL Survey questions with higher scores indicative of greater acceptability. Caregivers answered the following three questions: (1) "*How was your experience with delivery and pick up of the TRAVEL Study materials?*", (2) "*How was your experience getting set up for the eye tracking session (turning on the computer, connecting with your TRAVEL Study research assistant?*", and (3) "*How did you feel about the total length of time that the TRAVEL Study took to complete?*" Questions 1 and 2 were answered with a five-point Likert scale from “easy” (2) to “difficult” (-2). Question 3 was answered with three options: “too little time” (-1), “too much time” (-1), or “the right amount of time” (1).

Acceptability of eye tracking at home, specifically, was assessed by caregiver ratings of four questions designed to understand child focus, movement, toleration, and enjoyment during the session. Caregivers rated the following three questions on a three-point scale with answer choices of “yes,” “somewhat,” or “no”: (1) “*In your opinion, was your child able to focus during the eye tracking session?”,* (2) *“In your opinion, did your child move around a lot during the eye tracking session?”,* and (3) “*In your opinion, do you think your child tolerated the eye tracking session?”* Finally, caregivers rated a fourth question on a three-point scale with answer choices of “yes,” “neutral,” or “no”: (4) “*In your opinion, do you think your child enjoyed the eye tracking session?”*

2.4.1.3 Practicality. Practicality of TRAVEL was assessed by calculating four metrics. The first metric was the average length of study recordings with the designated start and end points of the participant logging onto and off of Zoom. The second metric was the percentage of children who were successfully calibrated for eye tracking and went on to complete the eye tracking portion of TRAVEL compared to those who failed calibration (as indicated by calibration stoppage on Tobii Pro Lab) and did not complete eye tracking. The third metric was the average calibration quality rated post-hoc with a screenshot of each child’s nine-point calibration results. Tobii Pro Lab labels each point on the viewing “plane” (the computer screen) according to the number of gaze samples provided by each child and points are labeled “Not enough data” if too few gaze samples are provided for that screen area. Calibration quality was scored as “Low” if 9 to 7 points are labeled as “Not enough data,” “Medium” with 6 to 4 points, and “High” with 3 to 0 points. See Figure 5 for an example of a participant calibration screenshot which would be rated as "Medium" in quality. The fourth practicality metric was the percentage of participants with usable eye tracking data (defined as at least 15 seconds of eye tracking data for at least one IJA and one RJA segment) relative to the total number of children who completed a study visit. We examined whether each metric differed between autistic and non-autistic children using independent samples t-tests.

2.4.2 Generation of TRAVEL Outcome Measures (Aim 2)

Each TRAVEL outcome measure was estimated using SEM methods. To plot the functional form of *TRAVEL trajectories* (H2a), we pulled data from fifteen-second video clips of Puppets and Robots. To form *RJA trajectories*, we used segments of *Puppets* in which children saw the examiner point to a puppet with a time delay (the RJA event). For *IJA trajectories,* we used segments of *Robots* in which children saw the robot beginning to spin next to the examiner (the IJA event). Since children saw the *Puppets* and *Robots* videos twice, with each video containing two trials (to the L and R of the examiner), trials with the "best" gaze data were chosen. For children who had sufficient data for each trial, the analyzed trial was chosen at random using a four-sided die.

Both fifteen-second clips were "binned" (segmented) by seconds and a proportion score per one-second bin were computed in which children's total time looking at the target (the puppet for RJA, the examiner for IJA) were divided by the total time looking at the full computer screen. For both *trajectories*, one-way ANOVAs were used to examine if child group, a time-invariant predictor (TIC), predicted initial differences in general screen looking at the start of each task (Bin 1). To examine the utility of *TRAVEL trajectories* in discriminating autistic and non-autistic groups (TICs), multiple group LCAs were leveraged and fit for each outcome according to child group. Multiple group LCAs allow for the assumption of population heterogeneity such that the model could accommodate differences in functional forms of outcome variables according to group. Each model was parameterized to account for nonlinear trajectories based on the following hypotheses.

We predicted that, based on the event type, that the functional form of JA would significantly differ according to TIC. We hypothesized that non-autistic children's *RJA trajectories* on average would take a bilinear spline functional form in which proportion scores of target looking (puppet) would dramatically spike and form a knot in response to the RJA event (examiner point) and would slowly decline thereafter (see Figure 6). Based on research suggesting that RJA is intact but still diminished with respect to referential understanding in autistic compared to non-autistic children (Congiu et al., 2016), we predicted that autistic children's *RJA trajectories* would similarly take the form of a spline regression line such that proportion scores of target looking would spike and form a knot in response to the RJA event and then decline thereafter, but that the spike would be significantly less dramatic than that of non-autistic children. A significantly smaller slope increase at the knot point would demonstrate that, on average, fewer autistic children responded to the RJA event than non-autistic children. In addition, the knots were modeled as random effects to account for the fact that target looking may have varied according to time and frequency per child.

For non-autistic children's *IJA trajectories*, we hypothesized that they would take the form of a spline regression line such that after the IJA event (the robot begins to spin), the proportion of scores of target looking (to the examiner for a "check in") would spike to form a knot and then decline thereafter (see Figure 7). We posited that the knot would occur a few milliseconds *after* the IJA event because we believed non-autistic children would look first at the activated robot and then to the examiner (the target). In contrast, based on research demonstrating consistent difficulties with IJA in autistic populations including diminished alternate gazing, we hypothesized that autistic *IJA trajectories* would take the form of a linear polynomial without a knot point. A linear IJA trajectory would signify that autistic children's proportion of *target* looking did not change in response to the IJA event, or that autistic children did not exhibit alternate gaze at the examiner after robot activation. Instead, since autistic children tend to exhibit more attention to moving objects, we hypothesized that their looking time would primarily be directed to the robot, so we posited that proportion scores of looking to the examiner by total screen looking would be low for the entirety of the IJA segment. Again, knots were modeled as random effects to account for the fact that target looking may have varied according to time and frequency per child.

To generate *total TRAVEL scores*, five IJA items (*Robots*, *Rabbit*, and *Balloon* proportion scores and *PCX* frequency scores (*Eye Contact* and *Gestures*) and three RJA items (*Story Time*, *Puppets*, and *RTN* proportion scores) were pooled and scores were estimated using exploratory factor analysis (EFA; Haig, 2005). Prior, correlations between items were examined in a correlation matrix with Pearson Correlation Coefficients and all items were examined for distributive properties. We then leveraged EFA using maximum likelihood (ML) estimation with the NLMINB optimizer in R to determine the adequate factor structure for IJA and RJA items first with an orthogonal (varimax) rotation and then with an oblique (quartimin) rotation. We also utilized functions of the PSYCH package in R including the fa command, which produces factor loadings based on a model's correlation matrices, and the fa.parallel command, which automatically generates a scree plot with the suggested number of factors appropriate for EFAs. We hypothesized that for *total TRAVEL scores*, a two-factor solution (IJA as one factor and RJA as the other) would best fit the data.

For EFAs and multiple LCAs, simulated critical values and eigenvalues were examined as indicators of model fit and factors were chosen using the Kaiser-Guttman rule which specifies that eigenvalues be retained if they are greater than one. Model fit was also established using root mean square error of approximation (RMSEA) and the comparative fit index (CFI), Tucker-Lewis index (TLI) with fit indices >.90 indicating good model fit for CFI and TLI, <.08 for SRMR, and <.05 for an excellent fit for RMSEA with a lower bound confidence interval that includes .10 (Brown, 2006).

To determine whether *total TRAVEL scores* could discriminate between groups, autistic and non-autistic group means were created for individual TRAVEL items and *total TRAVEL scores*. ANOVAs were conducted for each item. One-way ANOVA models were used for *Story Time* and *RTN* item proportion scores and *PCX* frequency scores. All other items, subdomains, and *total TRAVEL scores* were analyzed using repeated-measures ANOVA to examine group differences as a function of time, as children were shown these items twice and attention may have waned over time. In addition, receiver operator characteristic (ROC) curves were run for each item, subdomain (RJA and IJA), and *total TRAVEL scores*. ROC curves were used to determine how well items correctly discriminated between autistic and non-autistic groups, and area under the curve (AUC) measured the accuracy of this discrimination, ranging from 0.5 (differentiates no better than chance alone) to 1.0 (perfect differentiation) (Swets, 1988). AUC scores of > 0.5 indicate that a model is performing better than random guessing, scores of < 0.5 indicate a model is performing worse than random guessing, and scores > 0.7 are generally considered to have moderate discriminative validity (Swets, 1988). ROC curves provided measures of sensitivity, the “true positive” proportion of children who were correctly classified as autistic, and specificity, the “true negative” proportion of children correctly classified as non-autistic.

2.4.3 Construct Validity (Aim 3)

The construct validity of TRAVEL was examined accordingly. The convergent validity of the TRAVEL outcomes (*total TRAVEL scores* and *TRAVEL IJA* and *RJA trajectories*) were measured by examining associations with validated caregiver-report measures of (1) child JA, including social communication, social competence, and autistic characteristics via the SCQ total score and overall Total *T*-scores on SRS-2 (Hypothesis 3a), and (2) Expressive Vocabulary (MCDI-EV) and Receptive Vocabulary (MCDI-RV) scores on MCDI, developmental measures commonly associated with JA (Hypothesis 3b). The divergent validity of TRAVEL outcomes were examined by measuring their associations with the Restricted, Repetitive Behaviors and Interests factor score (SRS-RRB) on the SRS-2 since RRBs are thought to be theoretically unrelated to JA (H3c).

Convergent and divergent validity were examined with the following steps. Five Pairwise Partial Correlations were run for each TRAVEL outcome measuring its association with the five survey measures: (1) the SCQ, (2) the SRS-2, the (3) MCDI-EV and (4) MCDI-RV scores, and (5) the SRS-RRB score on the SRS-2. Modeled after analyses in Tenenbaum and colleagues (2021) and Murias and colleagues (2018), we controlled for participant age in months at the time of their study visit and cognitive development score on the MSD. Next, for each TRAVEL outcome that demonstrated a significant correlation with a survey measure, a hierarchical regression model was run in which the survey measure was regressed on the TRAVEL outcome. Each hierarchical regression model included three steps: step one included one of the five survey measures listed above; step two added child diagnostic status (autistic versus non-autistic); and step three, child biological sex. For correlations and regressions, we adjusted for multiple comparisons.

CHAPTER 3: RESULTS

3.1. Aim 1: Feasibility

3.1.1. Demand

Demand was defined as the probability of "program use" or the likelihood that individuals would participate in the TRAVEL Study. The total number of individuals who had knowledge of the TRAVEL study was 1,077, with 953 individuals who saw posts regarding the TRAVEL Study on Facebook and 124 from the Carolina Institute for Developmental Disabilities Autism Research Registry and Child Development Research Registry. There were 164 individuals total who inquired about the TRAVEL Study. All 164 individuals were contacted and of those, 50 were eligible and were scheduled. Out of the remaining 104, four declined to participate, four were ineligible due to location (outside of North Carolina) or age (too old), and the remaining (N=96) did not reply after the first inquiry or replied after recruitment was concluded. Of the 50 who were eligible and scheduled, two families (with autistic children) consented but did not participate and two new families with autistic children were subsequently recruited. Therefore, 50 families who were consented and eligible completed their visits, resulting in an attrition rate of 2/52 or 3.84%. Our total sample was 25 autistic children and 25 non-autistic children. The ratio of the total number of participating families who were eligible and who enrolled and consented (regardless of whether they completed the study) relative to the total number of families who had knowledge of the TRAVEL Study was 52: 1,077, or 4.82%.

3.1.2. Acceptability

Acceptability was defined as the extent to which the TRAVEL Study was suitable for the intended participants. A total acceptability score ranging between 0 to 12 was calculated for participating caregivers. Caregivers had an average total acceptability score of 9 out of 12 (*SD*= 3.33), indicating high acceptability of the TRAVEL Study. Each individual acceptability question was rated between 0 and 2 and average answers to each acceptability question are delineated in Table 3. Acceptability ratings were overall high, falling between 1.20 to 1.91 although two acceptability scores were low. Total length of time to complete the study received an average acceptability score of 0.66 (*SD*= 0.76) and an acceptability rating of child enjoyment of the eye tracking session received an average score of 0.74 (*SD*= 0.49).

3.1.3. Practicality

Practicality was defined as the extent to which the TRAVEL Study could be reasonably carried out with the intended participants using existing resources, means, and circumstances. TRAVEL practicality metrics were the length of study visits, percentage of participants successfully calibrated, average calibration quality, and percentage of participants with usable data.

The mean length of study visits was 36 minutes and 51 seconds (*SD*= 4 minutes, 36 seconds) and visit time did not significantly differ between the autistic and non-autistic groups, $t(45.13) = -0.62$, $p = 0.54$. All participants were successfully calibrated and went on to complete eye tracking. Two percent of participants had low calibration quality, 6.1% had medium quality, and 91.8% had high calibration quality. Calibration quality did not significantly differ between autistic and non-autistic groups, $t(30.45) = -1.19$, $p = 0.24$. One hundred percent of participants had useable data needed to generate *total TRAVEL scores*, 90% had useable data for *TRAVEL*

IJA trajectories, and 96% had useable data for *TRAVEL RJA trajectories*. Autistic and non-autistic groups did not significantly differ in rates of usable data acquired on binned data for Trajectory scores (*RJA trajectories*, $t(24)= 1.44$, $p= 0.16$, *IJA trajectories*, $t(47.32)= 0.40$, $p= 0.69$).

3.2. Aim 2: Generation of TRAVEL Outcome Measures

3.2.1 Generation of *TRAVEL Trajectories*

Multiple-group latent curve analyses (LCAs) were proposed to generate and examine JA trajectories. However, since children on average only exhibited target looking during one bin (the bin after the event, resulting in values of zero for all other bins), multiple-group LCA models run in SAS did not converge. Therefore, *IJA* and *RJA trajectories* per group were generated and examined using logistic regression. Each regression model includes a fixed knot point at the bin in which the event took place (Bin 4) and generates trajectories with an intercept term before and after each knot. Logistic regressions produced *RJA* and *IJA trajectories* nearly identical to those that would be generated with multiple group LCAs (Figures 8 and 9).

Trajectory intercepts and slopes tested for group differences in the log odds of looking to the target. Odds ratios (OR) with 95% confidence intervals (CI_{95}) are reported. For *RJA trajectories* (Figure 8), autistic and non-autistic children did not significantly differ in the odds of puppet looking onset after the examiner point (*RJA slope onset*, *OR*: 0.55, CI_{95} [0.26, 1.17], p=0.12) or puppet looking offset (*RJA slope offset*, *OR*: 0.90, CI_{95} [0.73, 1.00], p=0.34).

For *IJA trajectories* (Figure 9), autistic and non-autistic groups did not significantly differ in the odds of onset looking to the examiner eyes after the robot spun (*IJA slope onset*, *OR*: 1.01, CI_{95} [0.43, 2.36], p=0.96) but did significantly differ in the odds of offset looking to the examiner's eyes (*IJA slope offset*, 95% *OR*: 0.66, CI_{95} [0.45, 0.96], p=0.03). Autistic children had a significantly lower odds of looking to the examiner eyes after the robot spin, exhibiting steep slope decreases in target looking compared to non-autistic children. Non-autistic children had higher odds of maintaining target looking to the examiner after the event until the 15th bin and exhibited a less steep slope declination.

3.2.2 Generation of *Total TRAVEL Scores*

For *total TRAVEL scores*, IJA and RJA items (*Eye Contact*, *Gestures*, *Balloons*, *Puppets*, *Robots*, *Rabbit*, *RTN*, and *Story Time*) were pooled and scores were estimated using EFA with ML. A correlation matrix revealed high multicollinearity among many variables (see Figure 10), including high correlations between *Eye Contact* and *Gestures*, $r(47)= 0.74$, $p<.0001$, between *Puppets* and *Story Time*, $r(46)= 0.70$, $p<.0001$, and between *Story Time* and *Balloon*, $r(44)= 0.53$, $p<.0001$. Significant inter-correlations revealed that children who exhibited more of one behavior also exhibited more of other behaviors, making it difficult to estimate the unique effect of each JA behavior. Therefore, we were justified in using these eight variables to create a global JA score and test for variable reduction.

We used the fa.parallel and fa commands in the R PSYCH package to run our EFA. Fa.parallel suggested a two-factor structure would best fit the data and fa, that a two-factor structure with all eye tracking and *RTN* items (*Balloons*, *Puppets*, *Robots*, *Rabbit*, *RTN*, and *Story Time*) loaded onto one factor and *PCX* items (*Eye Contact*, *Gestures)* onto a second factor would best fit the data. The model had the following fit indices: $X^2(19)=27.54$, $p=0.09$, $CFI= 1.00$, $TLI= 0.94$, $RMSEA=0.07$, $CI_{90}= (0.00, 0.15)$, $SRMR=0.10$. The model had excellent fit for *CFI* and *TLI* and a good lower-bound confidence interval for *RMSEA*. *RMSEA* and *SRMR* values were close to the recommended values for good model fit (<0.05 for *RMSEA* and <0.80 for *SRMR*) at 0.07 and 0.10, respectively. Simulated critical values were, $X^2(19)=74.40$, $p<.0001$, suggesting evidence to reject the null hypothesis. A scree plot (see Figure 11) also suggested a two-factor structure: all eight variables loaded onto the first two components and the eigenvalues for both components was $\geq$1.

The model had two factors: one factor with eye tracking and *RTN* items (*Balloons, Puppets, Robots, Rabbit, RTN,* and *Story Time*) and a second factor with *PCX* items (*Eye Contact, Gestures*). Since *RTN* and eye tracking were structured attention tasks led by the examiner and caregiver in comparison to the *PCX*, in which JA was largely child-led, we hereafter refer to the first factor as "*adult-directed JA*" and the second, "*child-directed JA*."

3.2.3 Examination of TRAVEL Outcome Measures

Table 4 delineates descriptive statistics and ANOVAs of TRAVEL outcomes by child diagnostic status. Since *TRAVEL IJA trajectories* revealed that autistic and non-autistic children significantly differed according to slope declination, we include *IJA slope offset* scores as a TRAVEL outcome measure. Therefore, TRAVEL outcomes included five scores. Three scores were based on "*TRAVEL trajectories*": *IJA slope onset*, *IJA slope offset*, and *RJA slope onset*. Two "*total TRAVEL scores*" were based on our EFA: *adult-directed JA* factor scores and *child-directed JA* factor scores. ANOVAs demonstrated that groups did not differ on *TRAVEL trajectory* scores but did significantly differ on *total TRAVEL scores*: *adult-directed JA*, $F(1,47)= 4.00$, $p=0.05$; *child-directed JA*, $F(1,43)= 9.50$, $p=0.003$.

Next, ROC curves with *AUC* (area under the curve) measured how accurately TRAVEL outcome measures discriminated between autistic and non-autistic groups, ranging from 0.5 to 1 (see Table 5). TRAVEL Trajectory scores did not exhibit good discriminative accuracy. *IJA slope onsets* had discriminative accuracy worse than chance alone (*AUC*= 0.461) and *IJA slope offset* and *RJA slope onset* had relatively low accuracy (*IJA slope offset*, *AUC*= 0.571; *RJA slope* *onset*, *AUC*= 0.523), although *IJA slope offset* was slightly better. *total TRAVEL scores* exhibited moderate to good discriminative accuracy with an *AUC* of 0.704 for *child-directed JA* (Figure 12) and 0.743 for *adult-directed JA* (Figure 13). In other words, factor scores of *Eye Contact* and *Gestures* on the *PCX* (*child-directed JA*) had a 70% chance of discriminating between autistic and non-autistic children. Factor scores of all eye tracking tasks and *RTN* (*adult-directed JA*) had a 74% chance of discriminating between autistic and non-autistic children.

3.3. Aim 3: Construct Validity

For our third aim, we examined the construct validity of TRAVEL outcomes. We analyzed the construct validity of *IJA slope offset*, *adult-directed JA*, and *child-directed JA* because these measures significantly differed according to child groups and had moderate to good discriminative accuracy on ROC curves. We excluded *IJA slope onset* and *RJA slope onset* from validity analyses due to lack of group differentiation and poor discriminative accuracy. First we examined Pearson Partial Correlations for associations between TRAVEL Outcomes and survey measures. Second, survey measures were hierarchically regressed on each TRAVEL outcome in which step two added child diagnostic status (autistic versus non-autistic) and step three added child biological sex. Of note, biological sex did not contribute to the variance explained in any of our models. We adjusted for multiple comparisons in validity analyses due to the large number of statistical tests (15) by using a Bonferroni adjusted alpha level of 0.003 (0.05/15).

First, we analyzed <u>convergent validity</u> with our hypothesis that TRAVEL outcomes would be moderately to strongly associated with and significantly predictive of validated caregiver-report measures of JA by running correlations with TRAVEL outcomes and the SCQ Total Score and the SRS-2 Total *T*-Score (H3a). *Child-directed JA* was significantly negatively correlated with SCQ, *r*(43)= -0.36 , *p*= 0.01, and SRS-2 scores, *r*(43)= -0.47, *p*= 0.001 (see Figure 14), although only the latter correlation met the Bonferroni adjusted alpha level of 0.003. In other words, this finding illustrated that higher factor scores of *Eye Contact* and *Gestures* on the *PCX* (*child-directed JA*) were associated with lower SCQ and SRS-2 scores (higher scores indicate the presence of autism). *Adult-directed JA* had a negative association that approached significance with SCQ scores, *r*(45)= -0.28, *p*= 0.06, demonstrating a trend indicating that more JA behaviors exhibited on TRAVEL eye tracking tasks and *RTN* (*adult-directed JA*) were associated with lower scores on the SCQ. *IJA slope offset* did not exhibit significant associations with SCQ or SRS-2 scores.

Next, hierarchical regressions were run with TRAVEL outcomes and SCQ Total Scores and SRS-2 Total *T*-Scores. *Child-directed JA* was significantly negatively predictive of SCQ scores, *β*= -0.20, *p*= 0.04, although the Bonferroni correction was not met and *child-directed JA* no longer predicted SCQ scores with the addition of diagnostic status, *β*= -0.03, *p*= 0.66, or biological sex, *β*= -0.02, *p*= 0.75, (Table 6). This finding indicates that children with more *Eye Contact* and *Gestures* on the *PCX* (*child-directed JA*) had lower SCQ scores when controlling for age and cognitive development. However, *child-directed JA* was no longer predictive of SCQ scores when accounting for diagnostic status, which demonstrates that diagnostic status was strongly predictive of SCQ scores and suggests that *child-directed JA* was also strongly related to diagnostic status.

Child-directed JA also significantly negatively predicted SRS-2 scores, *β*= -0.60, *p*= 0.002, and met the Bonferroni correction. However, the relationship between *child-directed JA* and SRS-2 scores only approached significance with the addition of diagnostic status, *β*= -0.27, *p*= 0.07, and biological sex, *β*= -0.30, *p*= 0.06, (Table 7). In other words, children with more *Eye*

Contact and *Gestures* on the *PCX* (*child-directed JA*) had lower SRS scores when controlling for age and cognitive development. However, the relationship between *child-directed JA* and SRS-2 scores only trended towards significance when accounting for diagnostic status, which was strongly predictive of SRS-2 scores. Step two of this model accounted for 70% of the variance explained in children's SRS-2 scores (Adj $R^2 = 0.70$). *IJA slope offset* and *adult-directed JA* were not significantly predictive of SCQ or SRS-2 scores.

Next, we analyzed our hypothesis that TRAVEL outcomes would exhibit convergent validity in that they would be moderately to strongly associated with and significantly predictive of expressive (MCDI-EV) and receptive vocabulary (MCDI-RV) scores on the MCDI (H3b). Contrary to our hypothesis, TRAVEL outcomes exhibited no significant correlations with MCDI-EV or MCDI-RV scores nor were they significantly predictive of MCDI-EV or MCDI-RV scores in hierarchical regressions. We completed independent samples t-tests post hoc to ensure that our language measure differentiated between groups as expected (young autistic children show typically show lower overall language scores compared to non-autistic children, particularly on receptive language measures (Charman et al., 2003)). Non-autistic children had significantly higher MCDI-EV, $t(20)=4.13$, $p= 0.0005$, and MCDI-RV, $t(20)= 2.86$, $p= 0.009$, scores compared to autistic children as expected. However, average scores on the MCDI-EV (autistic median (*m*)= 67.00, interquartile range (*IQR*) (43.00, 91.00); non-autistic *m*= 87.00 (85.00, 89.00)) and MCDI-RV (autistic *m*= 81.00 (72.00, 89.00); non-autistic *m*= 87.00 (86.00, 88.00)) with a range of possible scores from 0 – 89 indicated that children in both groups exhibited ceiling effects.

Finally, we examined <u>divergent validity</u> by analyzing our hypothesis that TRAVEL outcomes would be weakly associated with and not significantly predictive of SRS-2 Restricted and Repetitive Behaviors and Interests *T*-Scores (SRS-RRBs) since research suggests that restricted and repetitive behaviors and interests are theoretically unrelated to JA (H3c).

In line with our hypotheses, *IJA slope offset* did not exhibit significant associations with SRS-RRB scores, $r(46)= 0.04$, $p= 0.76$. Contrary to our hypothesis, an association between *adult-directed JA* and SRS-RRB scores approached significance, $r(45)= -0.26$, $p= 0.06$, although hierarchical regressions demonstrated that neither *adult-directed JA* Scores, $\beta= -0.11$, $p= 0.33$, nor *IJA slope offset*, $\beta= -77.82$, $p= 0.32$, were significantly predictive of SRS-RRB scores. In other words, neither JA behaviors exhibited on eye tracking tasks and *RTN* (*adult-directed JA*) nor latency looking away from the examiner during *Robots* (*IJA slope offset*) influenced children's propensity for restricted and repetitive behaviors and interests, although a trend showed that more JA on eye tracking tasks and *RTN* (*adult-directed JA*) was associated with fewer restricted and repetitive behaviors and interests.

Results of the *child-directed JA* divergent validity model were contrary to our hypothesis that TRAVEL outcomes would be weakly associated with and not significantly predictive of SRS-RRB scores. *Child-directed JA* scores were significantly negatively correlated with SRS-RRB scores, $r(43)= -0.47$, $p= 0.001$, and negatively predictive of SRS-RRB scores, $\beta= -0.61$, $p= 0.002$, meeting the Bonferroni adjustment on both accounts. However, the relationship between *child-directed JA* and SRS-RRB scores only approached significance in step two when accounting for diagnostic status, $\beta= -0.30$, $p= 0.08$, but was significant again in step three of the model when accounting for diagnostic status *and* biological sex, $\beta= -0.304$ $p= 0.05$, (see Table 8). In other words, more *Eye Contact* and *Gestures* on the *PCX* (*child-directed JA*) was associated with and predictive of fewer restricted and repetitive behaviors and interests when controlling for age and cognitive development and over and above the effects of diagnostic status and biological sex. In addition, the model explained a large portion of the variance in SRS-RRB scores in step one with predictors alone (Adj R^2 = 0.41) and steps two when accounting for diagnostic status (Adj R^2 = 0.61) and three (Adj R^2 = 0.61) when accounting for diagnostic status and biological sex.

We examined correlations and hierarchical regressions with *child-directed JA* and the second factor score on the SRS-2, Social Communication and Interaction (SRS-SCI), post hoc to ensure that the convergent validity exhibited between *child-directed JA* and SRS Total *T*-Scores (H3a) was not due solely to strong associations between *child-directed JA* and the SRS-RRB factor score. *Child-directed JA* had a significant negative association with SRS-SCI scores, $r(43)= -3.42$, $p= 0.001$, and was significantly predictive of SRS-SCI scores, $\beta= 70.23$, $p= 0.0001$. Isolating the relationship between *child-directed JA* and SRS-SCI confirms that *child-directed JA* exhibited convergent validity with SRS Total *T*-Scores *along with* (unexpected) associations with SRS-RRB scores.

CHAPTER 4: DISCUSSION

There is a current dearth in remote neurodevelopmental assessment tools to examine child joint attention (JA) at home. The goal of the current study was to test the feasibility, utility, and validity of the TRAVEL Study, a remote and portable eye tracking and behavioral remote assessment of JA in young children aged 30 months to 54 months with and without an autism diagnosis. In line with our first hypothesis, we found that demand, acceptability, and practicality metrics proved TRAVEL to be a largely feasible remote assessment tool. Regarding our second hypothesis, both TRAVEL outcome measures differentiated autistic and non-autistic groups but not to the extent we hypothesized for *TRAVEL trajectories*. Autistic and non-autistic children did not significantly differ according to responding to joint attention (*RJA*) *trajectories.* Autistic children exhibited significantly different initiating joint attention (*IJA*) *trajectories* than non-autistic children although not according to functional form nor in *IJA slope onset* to target looking, but in *IJA slope offset*. Autistic children had significantly lower *total TRAVEL scores* than non-autistic children. Finally, in regards to our third hypothesis, the TRAVEL Study exhibited variable construct validity. Taken together, the TRAVEL Study brings a feasible remote assessment tool that can accurately differentiate between autistic and non-autistic children but that requires further investigation regarding the battery's validity. Henceforth we discuss our research findings in the context of the larger literature on remote assessment.

4.1 Feasibility

Feasibility metrics for the TRAVEL Study were overwhelmingly positive. Our results are in line with our hypothesis that there was sufficient demand for the TRAVEL Study to warrant further investigation. One hundred and sixty-four individuals inquired about the study, recruitment of 50 families was completed without issue, and attrition for the study was very low (3.84%). The number of individuals who received information about TRAVEL was high compared to the number who ultimately enrolled and consented (4.82%), which could be due to several factors. First, it is common for research studies to reach hundreds or even thousands of individuals through social media advertisements and see only a small percentage enroll (Aily et al., 2023). Aily et al. (2023) sought to understand recruitment yield for a telehealth intervention via social media advertisements and saw a steep drop-off from number of families reached through social media (measured via clicks) to the number actually enrolled. The study obtained a percentage of individuals knowledgeable of the study to individuals enrolled of 4.3%, near identical to our ratio of 4.82%, which suggests that our ratio was an appropriate measure of demand. Second, families with an autistic child may simply have less availability (to participate in research and overall) due to intensive service needs (Ellison et al., 2021). Finally, nearly 1,000 individuals knowledgeable about the study came from autism-specific groups on Facebook but it was not possible to determine how many of these individuals were (1) caregivers of autistic children (2) who *also* met the study's inclusion criteria. Therefore, although this demand statistic appeared low, it may not be reliable as a demand statistic alone.

According to caregivers and consistent with our hypothesis, the TRAVEL Study was considerably acceptable for families in order to measure JA in young children at home. These ratings indicated that caregivers saw study procedures as largely sufficient, especially pertaining to ease of material delivery and the ability for their children to engage in the eye tracking session. However, there were some areas of improvement noted. Caregiver ratings indicated that the eye tracking session could have been more enjoyable for children and that the timing of the study was too long. Although caregivers were informed of the post-visit survey portion of the study and its length (between 45 minutes to 1 hour and 15 minutes) prior to verbal consent, open feedback from participants indicated that study surveys took too much time. Since the average length of study visits outside of surveys was about 37 minutes and caregivers did not mention that the visit itself took too much time, future adaptations of the TRAVEL Study should attempt to shorten the survey requirements of the study to minimize research burden or provide additional compensation for survey time.

Several caregivers also stated that the eye tracking paradigms were not interesting enough for their children. However, useable eye tracking data rates were extremely high, suggesting that study stimuli were interesting enough to capture children's attention for the purposes of our research goals. Future adaptations of TRAVEL should inform caregivers that these tasks are designed to capture children's attention, but they will not be as diverting as other stimuli on screens like children's TV shows. Anecdotally, children's attention wavered most during *Story Time*, a much longer paradigm than others (over 3 minutes compared to others that totaled about 30 seconds each). Therein, to increase children's enjoyment during eye tracking (and already high rates of useable data), future adaptations of the TRAVEL Study should shorten *Story Time* significantly or offer several options for *Story Time* (i.e., multiple book options) to personalize the task to study children.

Finally, practicality metrics were in line with our hypothesis in that they indicated that TRAVEL could be reasonably carried out with the intended participants using existing resources, means, and circumstances. Visits were short (*m*=36.51 minutes, *SD*= 3.96) and lasted only a few minutes more than the hypothesized time (about 30 minutes). All children were successfully calibrated for eye tracking and calibration quality was high for nearly 92% of the sample.

Importantly, there were no significant differences in visit timing, calibration quality, or useable data between families with autistic and non-autistic children. This suggests that families with neurodiverse children were able to participate in a quick and accurate eye tracking study at home even if their children might have behaviors that would typically increase the length of a research visit or could affect the quality of eye tracking calibration, such as movement differences or reduced rates of overall attention. Alternatively, neurodiverse children may have been more comfortable, and potentially more compliant, at home, which has important implications for future eye tracking research on neurodiverse populations altogether. Useable data acquired was very high for *TRAVEL trajectories* (between 90-96%) and entirely complete (100%) for *total TRAVEL scores*. Ultimately, TRAVEL proved to be a practical means to acquire behavioral and portable eye tracking data on JA remotely in autistic and non-autistic children.

4.2 TRAVEL Outcome Measures

For Aim 2, we generated *TRAVEL trajectories* and *total TRAVEL scores* to test how two statistical methods – and JA assessment methods – could differentiate autistic and non-autistic children. Here we discuss the results of each outcome in the context of the prior literature. For *TRAVEL trajectories*, it is important to consider that proportion scores of looking to the target per bin were small, starting at 0 milliseconds and up for RJA and 200 milliseconds and up for IJA, which indicates that fixations to the target were brief across children. However, eye movement research shows that fixation lengths on average are typically between 200 to 350 milliseconds (Negi & Mitra, 2020) so this places children's fixations during trajectories within expected limits or longer.

First, we tested whether the functional form of RJA and IJA in response to an event differed according to child group by generating *TRAVEL trajectories* using logistic regression.

Contrary to our hypothesis that autistic children would exhibit diminished functional forms of RJA compared to non-autistic controls, or less overall time spent following the examiner point to a puppet, *TRAVEL trajectories* did not differentiate children according to RJA.

Lack of group differentiation in *RJA trajectories* could be explained by several factors including paradigm strength, the nature of RJA in autism, and the difference between gaze following and referential understanding. First, it is plausible that *RJA trajectories* did not differentiate between child groups because the eye tracking paradigm, *Puppets*, was not a strong measure of RJA. Indeed, children across groups only followed the examiner points about 60% of the time (see Table 2), *Puppets* scores did not significantly differ according to child group on an independent samples t-test, and useable data for proportion scores on *Puppets* was very high. Interestingly, *Puppets* was highly correlated with our other eye tracking measure of RJA, *Story Time* (r=.70, see Figure 10) and a group difference in *Story Time* scores approached significance. It could be that the puppet stimuli were just not interesting to children and the same measure of RJA using different, more interesting stimuli (perhaps *Robots*) should be tested.

Second, this finding could reflect literature suggesting that RJA alone may not be affected in autism to the extent of IJA (Mundy et al., 1986). Research is clear that autism is primarily defined by difficulties *initiating* rather than *responding to* social behavior (Mundy et al., 1990; Mundy & Crowson, 1997; Wetherby & Prutting, 1984) and it may be that *RJA trajectories* did not differentiate between child groups because a diagnosis of autism is not predictive of difficulties with RJA. Rather, research has shown that interactions between cognitive development and an autism diagnosis seem to best predict RJA in young children (Leekam et al., 1998; Mundy et al., 1994). Future research on TRAVEL should examine associations between child diagnosis, RJA, and the interactive effects of cognitive development.

Another possibility is that *Puppets* measured gaze following (i.e., social perception) rather than referential understanding (i.e., assigning meaning to entities referenced by others) which is essential for true RJA. Autistic children do not exhibit differences in gaze following compared to non-autistic children (Parsons et al., 2019). When tasks distinguish between gaze following and referential understanding, however, evidence suggests that RJA does indeed distinguish between non-autistic and autistic children (Congiu et al., 2016). One way to test this would be to add in an eye tracking paradigm similar to that created by Congiu et al. (2016) in which children are shown an object being hidden in one of two identical cups. The children are tasked with finding the object under two conditions, a gaze following condition in which the examiner gazes at the correct cup and the cups are visible, and a referential understanding condition, in which the cups are not visible, and the child has to rely solely on the examiner's gaze to select the correct cup. Adding in a task like so to future iterations of the TRAVEL Study may be a better way to understand the functional form of RJA and to measure RJA overall in autistic and non-autistic children.

We also tested whether *IJA trajectories* differentiated autistic and non-autistic children. Consistent with our hypothesis, autistic children exhibited significantly different functional forms of IJA compared to non-autistic children but not in the way we predicted. It was hypothesized that non-autistic children would exhibit a knot point indicative of a "check in" with the examiner after the IJA event, while autistic children would exhibit the opposite pattern. Rather, *IJA trajectories* did not differ according to *IJA slope onset*, or the latency of looking to the examiner's eyes after the robot spun. *IJA trajectories did* significantly differ in *IJA slope offset*, the latency of looking away from the examiner's eyes. In other words, both study groups had equal fixation times to the examiner's eyes for a "check in" after the robot began to spin.

Interestingly, the autistic group quickly looked away from the examiner's eyes – likely to look back at the robot – in that their proportion of time spent looking at the examiner's eyes dropped sharply. In contrast, the non-autistic group took significantly longer to disengage attention from the examiner's eyes than the autistic group.

Our *IJA trajectory* findings are contrary to our hypothesis that autistic children would not check in with the examiner at all after the robot began to spin. Children in both groups looked to the examiner at comparable rates immediately following the event. Instead, groups differed in disengaging from the examiner's eyes such that non-autistic children took significantly longer to disengage attention. Although this finding was unexpected, it is consistent with research showing that non-autistic children have attentional biases for faces. Non-autistic children show difficulty disengaging attention from faces compared to autistic children who do not show this bias and who disengage attention from faces more quickly (Chawarska et al., 2010). Research also generally equates longer fixations with deeper processing (Chawarska et al., 2016). This suggests that autistic children may be just as likely to attend to the eyes as non-autistic children but may show less processing of the eyes overall. Moreover, autistic children tend to show visual preferences for non-social stimuli, particularly *objects* in motion, as well as atypical and/or sensorial object exploration (Mottron et al., 2007; Ozonoff et al., 2008). It is therefore likely that autistic children quickly disengaged attention from the examiner's face in favor of looking to the robot. Future iterations of the TRAVEL Study should test this finding further by investigating *IJA trajectories* with a larger sample size and with other IJA eye tracking items such as *Balloon* or *Rabbit*.

For our second outcome measure, we used exploratory factor analysis (EFA) to generate an underlying latent JA variable – *total TRAVEL scores* – made up of several indices of JA on eye tracking and behavioral measures. We hypothesized that a two-factor solution, IJA as one factor and RJA as the other, would best fit the data. Although our EFA yielded a two-factor structure, the makeup of the factor structures was not as hypothesized. All eye tracking measures and response to name (*RTN*) loaded onto one factor (*adult-directed JA*) and parent-child interaction (*PCX*) items *Eye Contact* and *Gestures* loaded onto a second factor (*child-directed JA*), rather than overarching IJA and RJA factors. Each eye tracking item loaded onto the first factor which shows task consistency in eye tracking measures even though some items measured RJA and others, IJA. Interestingly, *RTN* was our only behavioral measure of RJA, and it loaded onto the first factor (*adult-directed JA*) with eye tracking items rather than the second factor (*child-directed JA*) with other behavioral items from the *PCX*.

We hypothesize that the loading of *RTN* and the overall factor structure were due to the nature of TRAVEL tasks. Even though *RTN* was a behavioral measure, it required a specific response and *attention* compared to other *PCX* items. Both *RTN* and eye tracking paradigms were arranged so that children were seated and directed to different attentional cues by an adult (the examiner or caregiver). The *PCX* was a largely unstructured free play task between the caregiver and child in which incidence of JA was likely child directed. From this overarching standpoint, it is possible that the factor structure of *total TRAVEL scores* may still be representative of RJA as one factor (*adult-directed JA*) and IJA as the other (*child-directed JA*). Eye tracking included items (*Robots, Rabbit, Balloon*) designed to measure IJA through alternate gaze (i.e., shifting attention between a person and object) but the "interaction" itself was still initiated by the examiner and children exhibited alternate gaze *in response to* the interaction. Future research would be well-served to examine the construct validity of each factor score with validated measures of IJA and RJA specifically to explore this possibility further.

Next, we tested whether *total TRAVEL scores* differentiated between child groups. We hypothesized that *total TRAVEL scores* would significantly differentiate children such that autistic children would have lower overall scores than non-autistic children (H2b). Consistent with our hypothesis, autistic and non-autistic children showed significant differences in *adult-directed JA* and *child-directed JA* scores such that non-autistic children exhibited eye tracking, *RTN*, and *Eye Contact* and *Gesture* scores on the *PCX* about twice the magnitude to those of autistic children (see Table 4). This finding demonstrates that *total TRAVEL scores* strongly differentiated child groups and aligns with literature showing that difficulties with JA across eye tracking and behavioral indices are a core feature of autism (Charman, 2003; Nyström et al., 2019).

ROC curves were an important qualifier for the utility of TRAVEL outcomes. *Adult-directed JA* and *child-directed JA* scores had receiver operator characteristic (ROC) curves with area under the curve (AUCs) of .74 and .70 indicative of good clinical utility in discriminating between autistic and non-autistic groups. *Total TRAVEL scores* far outperformed ROC curves of TRAVEL Trajectory scores. ROC curves for *IJA slope offset* and *RJA slope onset* had AUCs of .52 and .57, respectively. These AUCs signify that *IJA slope offset* and *RJA slope onset* only have a 52% and 57% chance of distinguishing between child group. *IJA slope onset* had an AUC of .46 indicative of discriminative validity worse than chance alone. Moreover, ROC curve analyses suggest that *TRAVEL trajectories* did not predict child diagnostic status even with marked group differences in *IJA slope offset*. ROC curve analyses further suggest that *total TRAVEL scores* are a good predictor of whether a child has an autism diagnosis and show promise for the utility of the TRAVEL Study as a whole.

4.3 Construct Validity

We analyzed construct validity in three of the five TRAVEL outcomes, *child-directed JA*, *adult-directed JA*, and *IJA slope offset*, due to their clinical utility in discriminating between groups. First, we hypothesized that outcomes would exhibit moderate to strong associations and relationships with validated caregiver-report measures of JA, the Social Communication Questionnaire (SCQ) Total Score and the Social Responsiveness Scale (SRS-2) Total *T*-Score (H3a). Overall, the three TRAVEL outcome measures had relatively poor convergent validity with survey measures of JA. Neither *adult-directed JA* nor *IJA slope offset* were significantly associated with or predictive of survey measures of JA. Only *child-directed JA* showed evidence of convergent validity in that it was significantly negatively predictive of SRS-2 scores even after accounting for age, cognitive development, diagnostic status and biological sex. Importantly, *child-directed JA* was not predictive of the SCQ after accounting for diagnostic status which calls into question the validity of *child-directed JA* since both the SRS-2 and SCQ are screening measures of JA.

What about *child-directed JA* was negatively related to SRS-2 scores but not SCQ scores above and beyond diagnostic status and biological sex? It could be that scores on the SRS-2 reflect other child factors also related to *child-directed JA* that are not picked up by the SCQ. The SCQ is a general measure of social communication and social skills. Alternatively, the Social Communication and Interaction factor score on the SRS-2 is made up of four specific social domains: Social Awareness, Social Cognition, Social Communication, and Social Motivation. Perhaps *child-directed JA* is related to one or a unique combination of these specific domains such as social motivation. Indeed, autistic females exhibit levels of social motivation similar to non-autistic females and because of this, tend to exhibit more social behaviors than do autistic males (Head et al., 2014; Sedgewick et al., 2016). Children's social motivation alone is sure to influence how they performed on items that make up *child-directed JA – Eye Contact* and *Gestures* on the *PCX*.

Another possibility is that *child-directed JA* was related to other child symptoms picked up by the SRS-2 and not the SCQ. For example, autistic children with co-occurring ADHD and/or co-occurring anxiety show compounded difficulties with JA in that they exhibit elevated scores on the SRS-2 (Factor et al., 2017). It is therefore possible that *child-directed JA* could be reflective of children who flag higher on the SRS-2 due to the presence of autism and ADHD. Future research should examine associations between *child-directed JA*, specific indices of social communication and interaction, and other co-occurring conditions that may yield elevated scores on JA screeners like ADHD to confirm validity.

Moreover, questions remain about convergent validity due to unexpected language findings. Contrary to our hypothesis that TRAVEL outcomes would be strongly associated with and significantly predictive of receptive and expressive vocabulary on the Mac-Arthur Bates Communicative Developmental Inventory (MCDI) (H3b), none were significantly associated with either MCDI scores. Decades of research have shown that JA and language are strongly related in infancy and early childhood (Markus et al., 2000; Tomasello & Todd, 1983) and RJA and IJA in the first two years of life are uniquely predictive of receptive and expressive language at ages two and three (Markus et al., 2000; Watt et al., 2006).

While unexpected, there are several possibilities that may explain this lack of relationship. It is plausible that a lack of association between TRAVEL outcomes and language was due to the limitations posed by a caregiver-reported language measure. In addition, it is possible that the choice of language measure was not the best fit here, as it is difficult to find a strong measure that is appropriate for a sample of young children in which half were autistic and half non-autistic. Unfortunately, an obstacle of conducting research remotely was our inability to assess child language with in-person behavioral measures typically used in pediatric research. We chose the MCDI because, to our knowledge, it was the only available caregiver-reported language measure for this age group with good psychometric properties. However, this measure resulted in very little variability due to ceiling effects for the non-autistic group and a majority of the autistic group, making it difficult to draw substantive conclusions about TRAVEL outcomes and their relationship with language on the MCDI. Future iterations of TRAVEL (and remote research studies in general) would be well-suited to measure child language with batteries such as the NIH Toolbox Picture Vocabulary Test (TPVT) (Gershon et al., 2014). The TPVT is a computerized assessment of language completed on a tablet that has a start point based on age and demographic factors and continually adapts to the child based on his or her performance. An iPad with the TPVT could easily be shipped in TRAVEL boxes and the TPVT is quite short and accessible. Caregivers could easily administer this program to their children for a more accurate measure of language, which would give us more sufficient means of examining convergent validity of TRAVEL outcomes.

Finally, divergent validity findings were also contrary to our hypotheses. We hypothesized that TRAVEL outcomes would be weakly associated with and not significantly predictive of Restricted, Repetitive Behaviors and Interests factor scores (SRS-RRBs) on the SRS-2. This hypothesis was based on research suggesting that JA is theoretically unrelated to RRBs (Mundy et al., 1994). Only *IJA slope offset* exhibited divergent validity with weak associations and no significant predictions with SRS-RRB scores on hierarchical regressions. However, we are cautious to draw substantive conclusions from this finding given that *IJA slope* *offset* had poor accuracy discriminating between child groups on ROC Curves (even though groups had significantly different *IJA slope offset* scores).

In contrast, *adult-directed JA* had a near significant association with SRS-RRBs (p= 0.06) and *child-directed JA* was significantly associated with and predictive of SRS-RRBs after accounting for age, cognitive development, diagnostic status, and biological sex. These findings are meaningful because both *adult-directed JA* and *child-directed JA* had relatively strong clinical utility discriminating between autistic and non-autistic groups. Significant associations and relationships between *total TRAVEL scores* and RRBs are contrary to a lack of theoretical relatedness found by Mundy et al. (1994). It may be that this is a slightly outdated hypothesis. After all, both JA and RRBs are necessary to make an autism diagnosis and indeed, more recent research has found that JA is related to RRBs to some degree. For example, studies have examined relationships between measures of social communication/interaction and of RRBs and found that the constructs exhibit correlations between 0.32 to 0.57 (Gotham et al., 2006; Lecavalier et al., 2020; Lord et al., 2000). Interestingly, the frequency in which children exhibit RRBs has shown to be unassociated to the "severity" of social communication difficulties measured on the ADOS-2 (Harrop et al., 2014). Therefore, it is important to take into account that JA (and social communication and interaction more broadly) and RRBs may be related to some degree but are nonetheless two distinct concepts that make up separate factor structures on diagnostic tools like the ADOS-2 (Lord et al., 2000). Future research should take the opportunity to understand the divergent validity of TRAVEL outcomes with other measures outside of the realm of constructs used in an autism diagnosis.

Taken together, the TRAVEL Study requires further study with larger sample sizes and changes to some survey and paradigm measures due to its questionable construct validity. *Child-*

directed JA seemed to be the strongest measure of convergent validity but was related to only one caregiver-report measure of JA (the SRS-2 but not the SCQ), unrelated to language, and exhibited divergent validity findings near the inverse of our hypotheses. Future research should examine how *child-directed JA* could be differentially related to measures of JA on the SRS-2 but not the SCQ. Further research should also measure the convergent validity of TRAVEL outcomes in relation to better, more adaptive measures of child language. Importantly, three TRAVEL outcomes were significantly discriminative of child groups on ROC Curves which indicates that TRAVEL would nonetheless be a useful research tool in identifying features of autistic and non-autistic development from home.

4.4 Strengths and Limitations

The current study had several strengths and limitations to consider. The use of portable eye tracking itself is both a strength and limitation of this study. Although portable eye trackers are significantly more cost effective than stationary eye trackers, they can still be costly (in the thousands). In addition, any technology is subject to damage and sending materials to families with young children can increase that risk. However, benefits of portable eye tracking still outweigh many cons. Portable eye tracking is a highly acute, objective, spectral measure of attention that can be done outside of the lab. Portable eye tracking can mitigate the risk of implicit bias often seen in in-person neurodevelopmental assessment. Moreover, eye tracking at home is more flexible, comfortable, and naturalistic for study children without the presence of an unfamiliar examiner and environment. Finally, eye trackers may lower or negate costs of other research expenses such as labor costs for behavioral coding.

Both a strength and limitation of TRAVEL is the completion of eye tracking at home. On the one hand, children were likely more comfortable completing the experiment at home with just their caregivers. However, study protocol was often impacted without an examiner and/or study team present to help caregivers with behavioral support and to mitigate distractions. For example, caregivers were instructed to seat their child in a chair with a strap (if delivered study supplies) or in a car seat or their lap (if mailed supplies). However, three caregivers reported that their child would become upset if seated and each were out of their seat for at least some portion of the session. These three children each had poorer data quality on many tasks and were the participants with the majority of missing data. These children may have been seated successfully if an examiner familiar with eye tracking children was present to help.

Caregivers were also advised to have childcare arranged for any siblings and to put pets away during the experiment. However, we recognize that this was a very tall order, and many caregivers were not able to secure care for others within their home during the study visit. Therefore, it was a common occurrence for siblings and pets to want to join in on the experiment. Another disadvantage of eye tracking at home is that lab settings often have playrooms and childcare available. Finally, it was impossible to mitigate other distractions without control of the environment, unlike a lab setting. Children were often distracted by toys from home and technologies like TVs and iPads. Adults could not control if their doorbell or phone wrang. Nevertheless, TRAVEL had extremely high useable data rates and caregivers had high acceptance of the study which speaks to the powerful nature of remote research in its flexibility and convenience.

A major limitation of the study was that Tobii products are not made for *remote* eye tracking even if they are made for *portable* eye tracking. Eye trackers are typically configured with two computer screens in lab and portable eye tracking settings. The participant will view the stimuli on one screen. On the second screen the data collector uses the administration page to control the experiment, monitor the participant with live video, and videorecord the participant for a behavioral record of the experiment for later viewing and data corroboration. This setup was impossible for remote eye tracking. Since we were not in-person during data collection and controlled the singular participant laptop on Zoom, we were not able to pull up the administration page on our examiner laptop. We calibrated the participant and began the experiment using the administration page. However, with just a singular participant computer, the administration page disappears after the experiment begins so that participants can be eye tracked. Tobii Lab Pro continues to videorecord the participant offline during the experiment, so we were able to obtain a behavioral record of the participant post hoc. However, we were unable to monitor participants live during the experiment without access to the administration page.

Tobii Pro Lab and Zoom programs are also not yet fully compatible and this is an important element that needs to be addressed for future remote eye tracking studies. Participants were videorecorded on Zoom during the entirety of the visit so we could record *RTN* and *PCX* tasks. Although we were unable to monitor participants live without the Tobii administration page, we originally thought we could have monitored participants live on Zoom. However, Tobii Pro Lab and Zoom are not yet able to videorecord simultaneously. To videorecord the participant during eye tracking with Tobii Pro Lab, participants were not able to be on their Zoom camera. Therefore, it was up to caregivers to monitor their children during the eye tracking task. Lack of live monitoring complicated data collection often and had possible implications for data reliability.

Another major limitation of the study as it pertains to mailing supplies was our reliance on shipping companies. Many visits had to be rescheduled due to tight timelines between participant visits, only one full set of TRAVEL Box materials, and shipments not arriving at their scheduled time. Anecdotally, shipments were delayed due to severe weather (FedEx was unable to leave the study box on properties if they were not covered and protected from the elements by an overhang), issues with participant addresses (namely, if their house number was not clearly displayed outside of their house), and participants living in extremely rural housing (in the Appalachian Mountains). Even with protective packaging, materials were also subject to and did sustain damage. The participant laptop broke during one shipment (when about two-thirds of recruitment was complete). The laptop took about six weeks to repair, and this delay forced us to reschedule six participants. Unfortunately, this influenced attrition when one family withdrew from the study during rescheduling. Future versions of the TRAVEL Study should acquire complete materials for several TRAVEL Box sets. Several Box sets could offset problems with shipping that cannot be controlled. Finally, with many things to keep track of, caregivers sometimes forgot to place their TRAVEL Boxes at their front door during their scheduled FedEx pick-up windows. Shipments were therefore often delayed. We found that the best way to control for this was to schedule pick-ups immediately after study visits so that caregivers prepared their boxes for reshipment right away.

Finally, a major limitation to consider is the homogeneity of the study sample. The sample was largely white non-Hispanic, autistic children were majority male, and caregivers had high, employment, annual household incomes, and levels of education. While the long-term goal of this research is to reach populations facing barriers to neurodevelopmental assessment and diagnostic delays, these families were not well represented in the current study and their representation is required in future iterations of this work. Children within this sample were also very young at 30 to 54 months (*m*= 47.0 months) meaning that autistic children in the sample were diagnosed earlier than the median diagnostic age of 53 months in the US (Maenner et al., 2023). This is likely due to the fact that children historically tend to be diagnosed earlier if they come from advantaged backgrounds (e.g., white and non-Hispanic) (Mandell et al., 2007, 2009; Shattuck et al., 2009). Finally, it is probable that statistical analyses were limited due to the study's small sample size. Autistic and non-autistic groups each contained only 25 children and it is generally recommended that statistical methods such as SEM have at minimum 30 cases per group (Wolf et al., 2013). Therefore, it is likely that our analyses were underpowered and future research with larger sample sizes is needed to make substantive conclusions regarding our study models.

4.5 Conclusion

In summary, this research project piloted the TRAVEL Study, a battery of behavioral and remote eye tracking tasks that can be completed with children at home, in 50 families with children aged 30 to 54 months. TRAVEL was a largely feasible means of remote eye tracking assessment for caregivers and their young children with and without an autism diagnosis. Several TRAVEL outcomes significantly discriminated between autistic and non-autistic children and scores generated using exploratory factor analysis had particularly good clinical utility. Nonetheless, TRAVEL outcomes did not exhibit strong construct validity and future iterations of the study are needed. This study has important implications for increasing the diversity of pediatric research studies altogether by bringing the eye tracking lab home to families.

Table 1. Demographic Variables for Study Children and Their Caregivers.

		Children	Caregivers[a]
Assigned Sex (*N, %)*	Female	20 (40)[b]	45 (93.9)
	Male	30 (60)	3 (6.1)
Ethnicity (*N, %)*	Hispanic/Latinx	5 (10.2)[c]	4 (8.3)
	Non-Hispanic/Latinx	44 (91.6)	44 (91.6)
Race (*N, %)*	White	42 (85.7)	41 (83.7)
	Black/African American	2 (4.1)	2 (6.1)
	Asian	1 (2.0)	2 (4.1)
	Multiracial	3 (6.0)	1 (2.0)
	Other Race	0 (0.0)	1 (2.0)
	Prefer not to answer	1 (2.0)	1 (2.0)
Age in Months (*M, SD*)		47.0 (6.86)	-
Highest Education Level (*N, %)*	≤ High School Degree	-	1 (2.1)
	Some College	-	1 (2.1)
	Bachelor's/Associate's Degree+	-	17 (35.4)
	Graduate Degree	-	29 (60.4)
Employment (*N, %)*	Employed Full- or Part-Time	-	40 (83.3)
	Unemployed	-	8 (16.6)
Annual Household Income (*N, %)*	$15,0001-$25,000	-	1 (2.1)
	$25,001 - $35,000	-	2 (4.2)
	$35,001 - $50,000	-	2 (4.2)
	$50,001 - $75,000	-	5 (10.4)
	$75,001 - $100,000 +	-	35 (73.0)

Note.

[a]Although we have demographic information from all study children (obtained during verbal consent/scheduling), only N=48 caregivers completed a demographic survey about themselves.

[b]There were equal numbers of males and females in the non-autistic group, but only 30% of autistic children were female.

[c]Ethnicity data was not obtained for one child.

Table 2. Descriptive Statistics and ANOVAs of Individual TRAVEL Items

Travel Items	**Autistic children** (mean, SD)	**Neurotypical children** (mean, SD)	***F***	***df***	***p***
Balloon	1.83 (1.70)	2.65 (1.23)	3.57	44	0.06
Eye Contact	6.68 (7.44)	10.68 (14.08)	1.57	48	0.21
Gestures	3.88 (4.50)	9.04 (5.76)	12.15	47	**0.001**
Puppets	0.62 (0.31)	0.65 (0.26)	0.11	46	0.74
Rabbit	1.74 (3.14)	2.57 (3.23)	0.77	44	0.38
RTN	0.48 (0.34)	0.82 (0.25)	15.99	48	**0.0002**
Robots	3.52 (5.08)	9.83 (7.72)	11.53	47	**0.001**
Story Time	0.68 (0.35)	0.83 (0.23)	3.077	48	0.08

Table 3. Acceptability Items

General Acceptability	M (SD)
1. How was your experience with delivery and pick up of the TRAVEL Study materials?	1.91 (0.35)
2. How was your experience getting set up for the eye tracking session (turning on the computer, connecting with your TRAVEL Study research assistant?	1.57 (0.68)
3. How did you feel about the total length of time that the TRAVEL Study took to complete?	0.66 (0.76)

Acceptability: Eye Tracking	M (SD)
1. In your opinion, was your child able to focus during the eye tracking session?	1.76 (0.52)
2. In your opinion, did your child move around a lot during the eye tracking session?	1.20 (0.83)
3. In your opinion, do you think your child tolerated the eye tracking session?	1.83 (0.38)
4. In your opinion, do you think your child enjoyed the eye tracking session?	0.74 (0.49)

Table 4. Descriptive Statistics and ANOVAs of Individual TRAVEL Outcome Measures

Travel Outcome Measures	**Autistic children** (mean, SD)	**Non-Autistic children** (mean, SD)	***F***	***df***	***p***
TRAVEL Trajectories					
IJA Slope Onset	0.02 (0.04)	0.01 (0.02)	0.61	41	0.43
IJA Slope Offset	-0.01 (0.02)	-0.02 (0.03)	0.84	41	0.36
RJA Slope Onset	0.01 (0.02)	0.01 (0.02)	0.11	46	0.73
Total TRAVEL Scores					
Adult-Directed JA	10.79 (10.62)	19.72 (19.23)	4.00	47	**0.05**
Child-Directed JA	9.56 (6.83)	17.78 (10.56)	9.5	43	**0.003**

Table 5. ROC Curves of Individual TRAVEL Outcome Measures

Travel Outcome Measures	**AUC)**	**95% CIs**
TRAVEL Trajectories		
IJA Slope Onset	0.461	(0.303-0.619)
IJA Slope Offset	0.571	(0.418-0.725)
RJA Slope Onset	0.523	(0.363-0.684)
Total TRAVEL Scores		
Adult-Directed JA	0.743	(0.600-0.887)
Child-Directed JA	0.704	(0.557-0.852)

Table 6. Child-Directed JA Hierarchically Regressed on SCQ Total Scores

Model	Variables	*β*	*SE*	*t*	*p*	R^2
1	Child-Directed JA	-0.20	0.09	-2.07	0.04	0.39
	Age in Months	0.08	0.14	0.57	0.56	
	Cognitive Development	-0.36	0.07	-4.59	0.0001	
2	Child-Directed JA	-0.03	0.08	-0.43	0.66	0.62
	Age in Months	-0.008	0.11	-0.07	0.94	
	Cognitive Development	-0.16	0.07	-2.26	0.02	
	Diagnostic Status	10.01	1.97	5.06	0.0001	
3	Child-Directed JA	-0.02	0.08	-0.31	0.75	0.61
	Age in Months	-0.003	0.11	-0.03	0.97	
	Cognitive Development	-0.15	0.07	-2.07	0.04	
	Diagnostic Status	10.03	1.99	5.02	0.0001	
	Biological Sex	-0.75	1.63	-0.46	0.64	

Table 7. Child-Directed JA Hierarchically Regressed on SRS Total *T*-Scores.

Model	Variables	*β*	*SE*	*t*	*p*	R^2
1	Child-Directed JA	-0.60	0.18	-3.29	0.002	0.47
	Age in Months	0.24	0.26	0.93	0.35	
	Cognitive Development	-0.68	0.14	-4.68	0.0001	
2	Child-Directed JA	-0.27	0.15	-1.80	0.07	0.70
	Age in Months	0.07	0.20	0.34	0.73	
	Cognitive Development	-0.29	0.13	-2.28	0.02	
	Diagnostic Status	19.72	3.51	5.61	0.0001	
3	Child-Directed JA	-0.30	0.15	-1.92	0.06	0.69
	Age in Months	0.05	0.20	0.28	0.77	
	Cognitive Development	-0.32	0.13	-2.39	0.02	
	Diagnostic Status	19.66	3.52	5.57	0.0001	
	Biological Sex	2.21	2.89	0.76	0.44	

Table 8. Child-Directed JA Hierarchically Regressed on SRS-RRBs *T*-Scores.

Model	Variables	*β*	*SE*	*t*	*p*	R^2
1	Child-Directed JA	-0.61	0.19	-3.18	0.002	0.41
	Age in Months	0.15	0.27	0.53	0.59	
	Cognitive Development	-0.62	0.15	-4.05	0.0002	
2	Child-Directed JA	-0.30	0.17	-1.76	0.08	0.61
	Age in Months	-0.01	0.22	-0.08	0.93	
	Cognitive Development	-0.25	0.14	-1.74	0.08	
	Diagnostic Status	18.71	3.96	4.71	0.0001	
3	Child-Directed JA	-0.34	0.17	-1.96	0.05	0.61
	Age in Months	-0.03	0.23	-0.16	0.87	
	Cognitive Development	-0.29	0.15	-1.93	0.06	
	Diagnostic Status	18.62	3.96	4.70	0.0001	
	Biological Sex	3.38	3.24	1.04	0.30	

Figure 1. Behavioral Measurement of Joint Attention

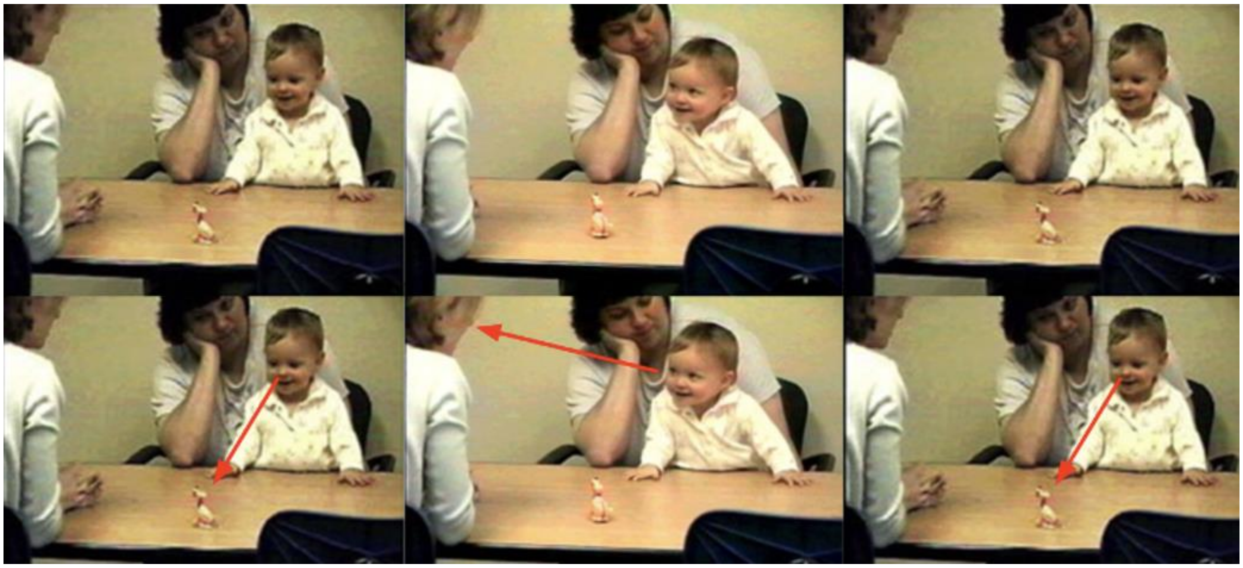


Figure 1. Images commonly used to visualize initiating joint attention (IJA) and alternate gaze by Mundy and colleagues (2003). Alternate gaze is an IJA behavior in which children switch their gaze from an interactive partner to an object, entity, or event with the goal of sharing the experience with that person for social learning or enjoyment.

Figure 2. TRAVEL Eye Tracking Tasks

Figure 2. TRAVEL eye tracking tasks designed to elicit initiating (IJA) and responding to joint attention (RJA). Clockwise from left are tasks *Balloons*, *Puppets*, *Story Time, Rabbit*, and *Robots.*

Figure 3. Timeline of TRAVEL Eye Tracking Tasks

VIDEO	STORY TIME	ROBOTS	PUPPETS	RABBIT	BALLOON	ROBOTS	PUPPETS	RABBIT	BALLOON
CONSTRUCT	WARM UP & RJA	IJA	RJA	IJA	IJA	IJA	RJA	IJA	IJA
TIME	3 MIN	35 SEC	1 MIN	35 SEC	45 SEC	35 SEC	1 MIN	35 SEC	45 SEC

Figure 3. Figure 3 shows a timeline of TRAVEL eye tracking tasks. *Story Time* was a warm-up activity meant to engage children and served as one measure of responding to joint attention (RJA). Children viewed four remaining tasks in which the examiner was seated between two items to her left and right. These tasks were shown and then repeated in the same consecutive order to account for data loss due to movement or inattention, but from right to left for the second set.

Figure 4. Response to Name (RTN) Procedures

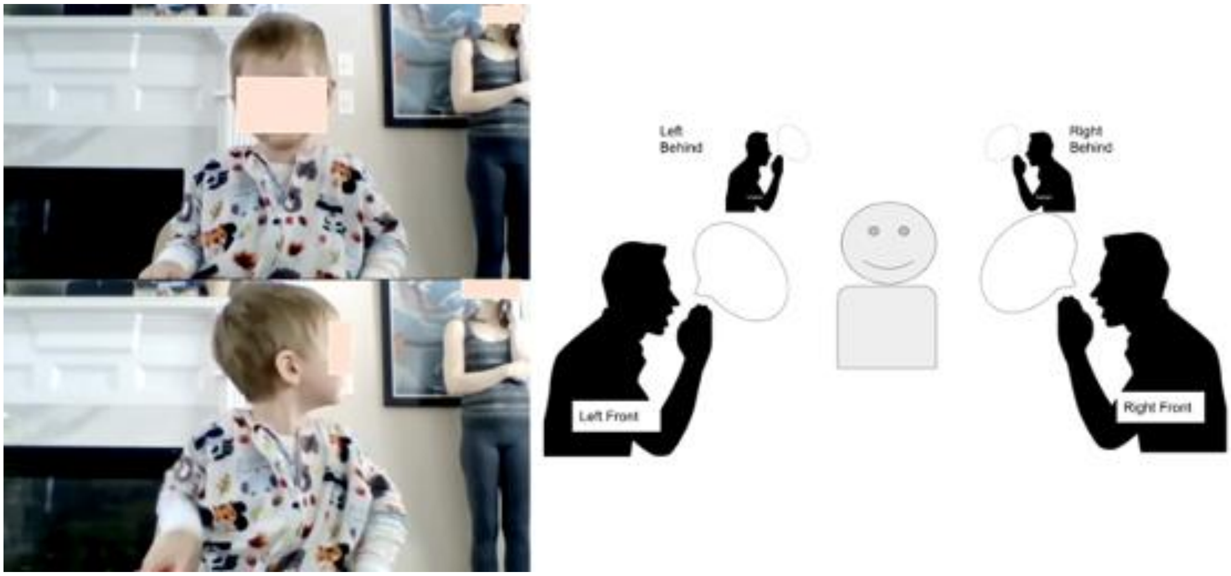


Figure 4. Response to name (*RTN*) procedures are shown in Figure 4. *RTN* is a form of responding to joint attention (RJA). Children were seated in front of the computer screen and were given a “distractor” toy (children were allowed to choose between a vibrating squirrel toy, a red fire truck, or both with the goal that children would choose the toy in which they were most interested). Caregivers were instructed to stand to the front left, front right, behind left, and behind right of their child (see Figure 4) and at each position, to call their child’s name twice loudly with a five second pause in between name calls. Videos were coded behaviorally and *RTN* proportion scores were calculated in which the number of “correct” responses to name (defined as a head turn *and* eye contact toward the caregiver after each name call) were divided by the total number of name calls.

Figure 5. An Example of Calibration Quality on Tobii Pro Lab

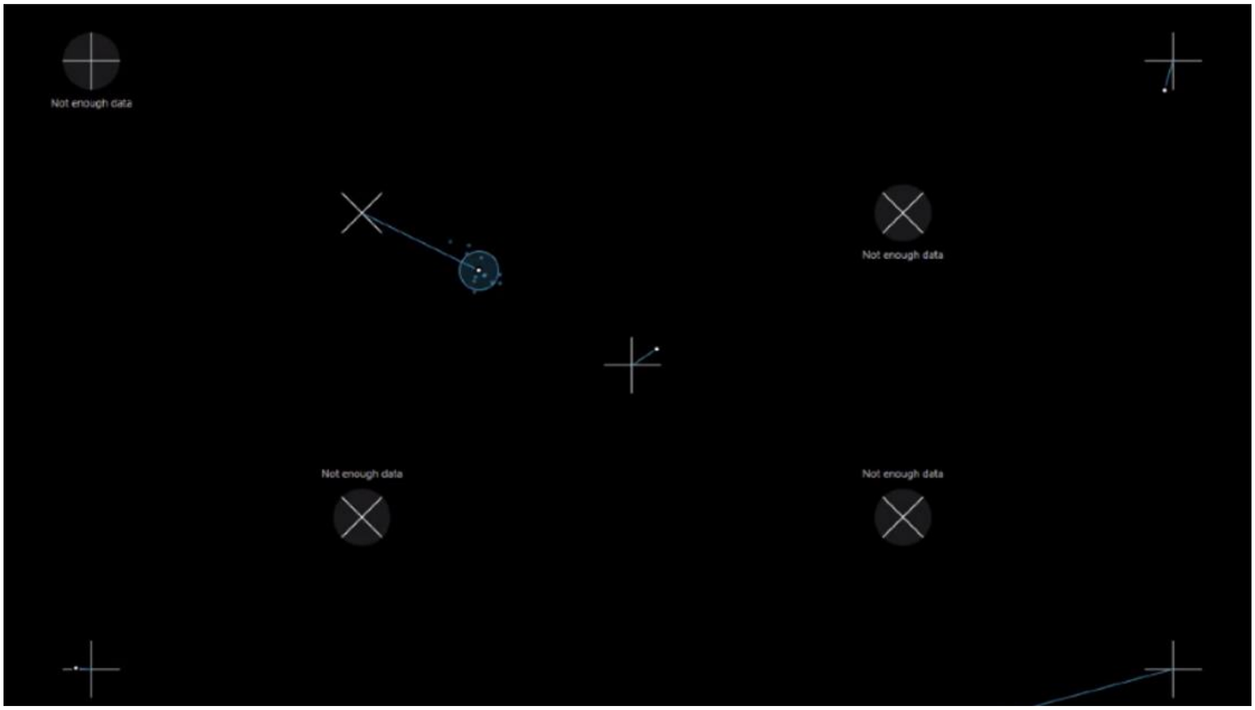


Figure 5. Average calibration quality was used as a metric of practicality in Aim 1 and was rated post-hoc with a screenshot of each child's nine-point calibration results. Tobii Pro Lab labels each point on the viewing "plane" (the computer screen) according to the number of gaze samples provided by each child and points are labeled "Not enough data" if too few gaze samples are provided for that screen area. Calibration quality will be scored as "Low" if 9 to 7 points are labeled as "Not enough data," "Medium" with 6 to 4 points, and "High" with 3 to 0 points. Figure 5 is an example of a participant calibration screenshot which would be rated as "Medium" in quality.

Figure 6. Hypothesized RJA Trajectories Per Autistic and Non-Autistic Groups

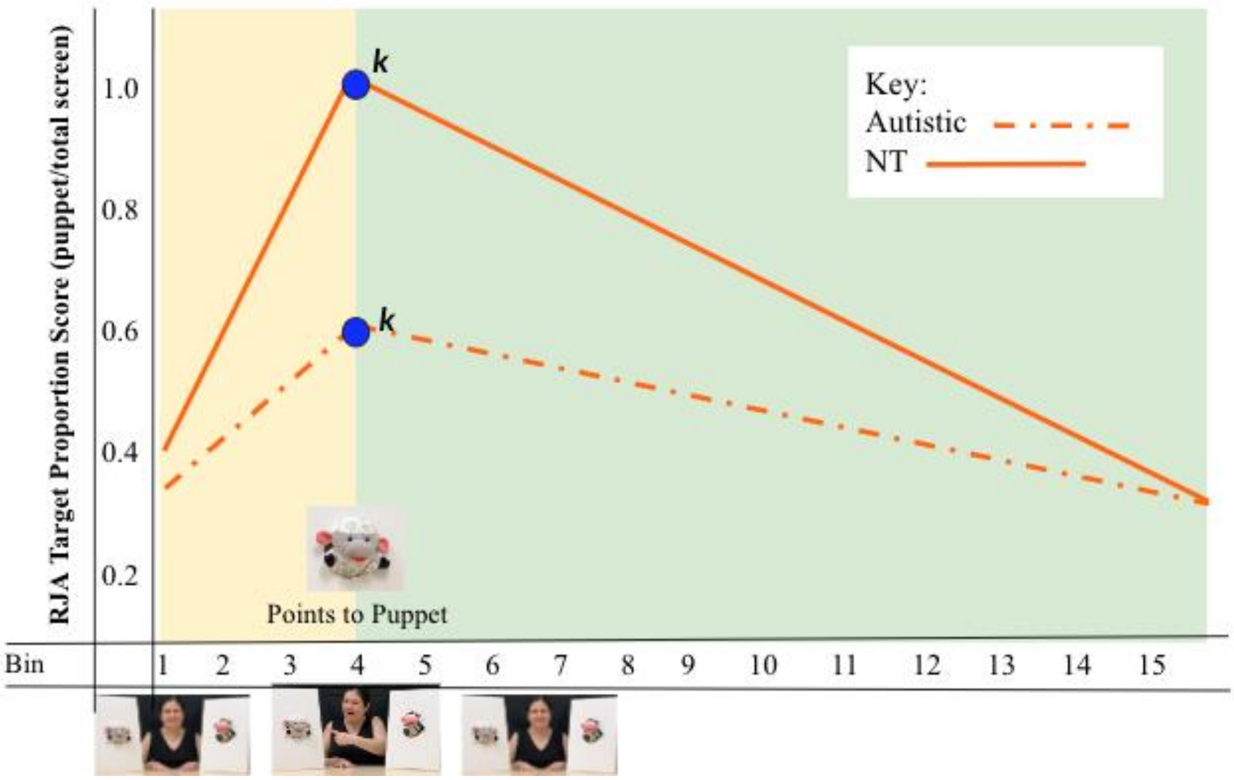


Figure 6. Figure 6 shows our hypothesized responding to joint attention (RJA) trajectories. We hypothesized that non-autistic children's *RJA trajectories* would take a bilinear spline functional form in which proportion scores of target looking (puppet) would dramatically spike and form a knot in response to the RJA event (examiner point) and would slowly decline thereafter. Based on research suggesting that RJA is intact but still diminished with respect to referential understanding in autistic compared to non-autistic children, we predicted that autistic children's *RJA trajectories* would similarly take the form of a spline regression line such that proportion scores of target looking would spike and form a knot in response to the RJA event and then decline thereafter, but that the spike would be significantly less dramatic than that of non-autistic children. A significantly smaller slope increase at the knot point would demonstrate that, on average, fewer autistic children responded to the RJA event than non-autistic children.

Figure 7. Hypothesized IJA Trajectories Per Autistic and Non-Autistic Groups

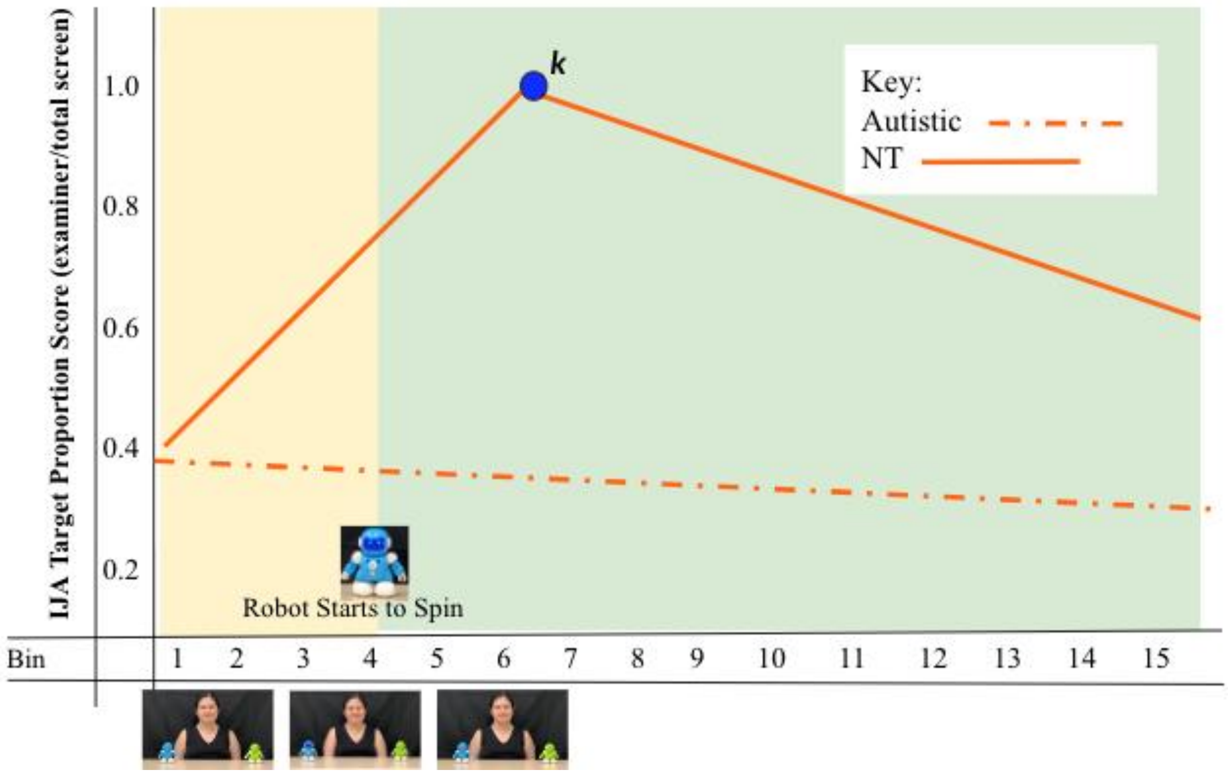


Figure 7. For non-autistic children's initiating joint attention (*IJA*) *trajectories,* we hypothesized that they would take the form of a spline regression line such that after the IJA event (the robot begins to spin), the proportion of scores of target looking (to the examiner for a "check in") would spike to form a knot and then decline thereafter. In contrast, based on research demonstrating consistent difficulties with IJA in autistic populations including diminished alternate gazing, we hypothesized that autistic *IJA trajectories* would take the form of a linear polynomial that would not exhibit a knot point. A linear IJA trajectory would signify that autistic children's proportion of *target* looking did not change in response to the IJA event, or that autistic children did not exhibit alternate gaze at the examiner after robot activation.

Figure 8. Logistic Regression: Responding to Joint Attention Trajectory

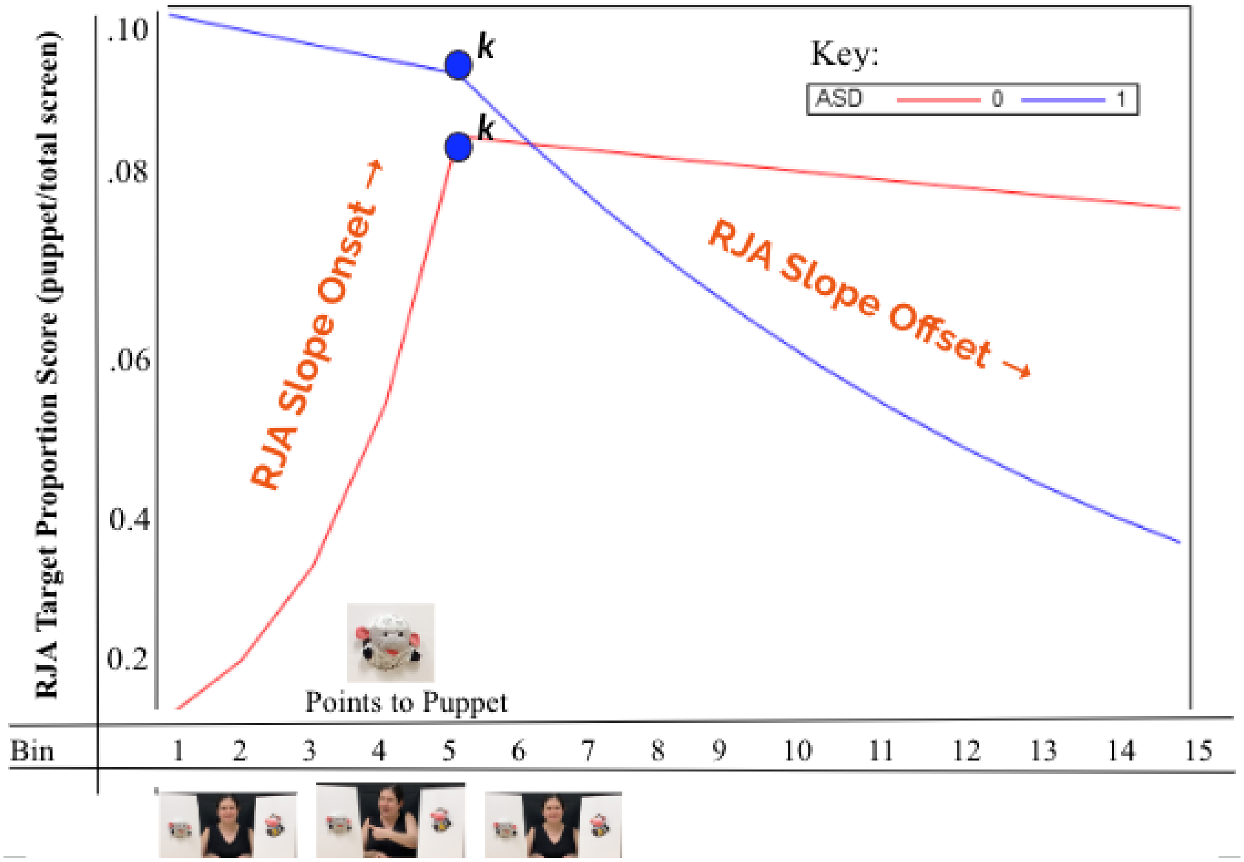


Figure 8. For responding to joint attention (*RJA*) *trajectories*, logistic regressions demonstrated that autistic and non-autistic children did not significantly differ in the odds of puppet looking onset after the examiner point (*RJA slope onset*, *OR*: 0.55, CI_{95} [0.26, 1.17], p=0.12) or puppet looking offset (*RJA slope offset*, *OR*: 0.90, CI_{95} [0.73, 1.00], p=0.34).

Figure 9. Logistic Regression: Initiating Joint Attention Trajectory

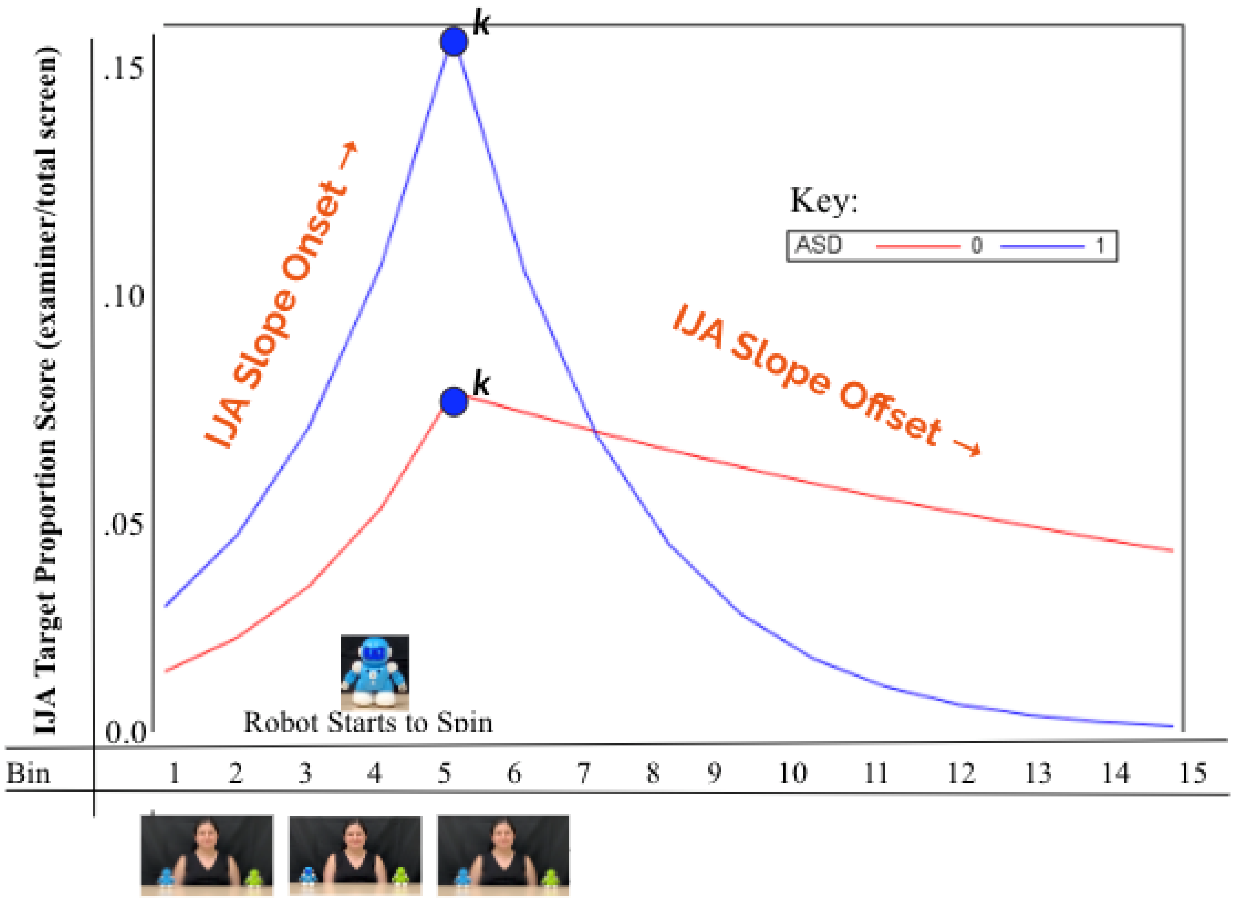
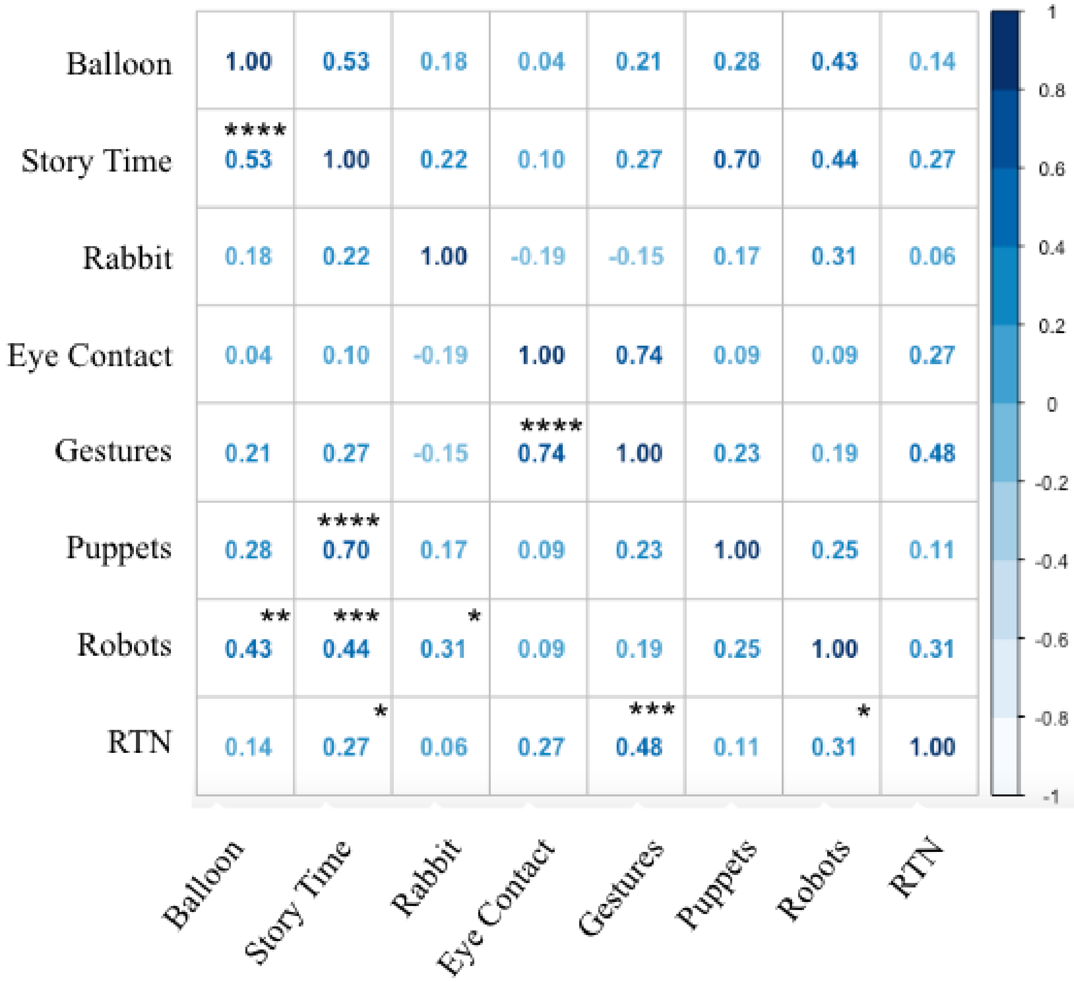


Figure 9. For initiating joint attention (*IJA*) *trajectories*, logistic regressions demonstrated that autistic and non-autistic groups did not significantly differ in the odds of onset looking to the examiner eyes after the robot spun (*IJA slope onset*, *OR*: 1.01, CI_{95} [0.43, 2.36], p=0.96) but did significantly differ in the odds of offset looking to the examiner's eyes (*IJA slope offset*, 95% *OR*: 0.66, CI_{95} [0.45, 0.96], p=0.03). Autistic children had a significantly lower odds of looking to the examiner eyes after the robot spin, exhibiting steep slope decreases in target looking compared to non-autistic children. Non-autistic children had higher odds of maintaining target looking to the examiner after the event until the 15th bin and exhibited a less steep slope declination.

Figure 10. Correlation Matrix: Total TRAVEL Score Variables



Note: *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$

Figure 10. A correlation matrix revealed high multicollinearity among many of the eight behavioral and eye tracking variables used for *Total TRAVEL scores*. Significant inter-correlations demonstrated that children who exhibited more of one behavior also exhibited more of other behaviors, making it difficult to estimate the unique effect of each joint attention (JA) behavior. Therefore, we were justified in using these eight variables to create a global JA score using exploratory factor analysis and to test for variable reduction.

Figure 11. Parallel Analysis Scree Plot for Total TRAVEL Scores

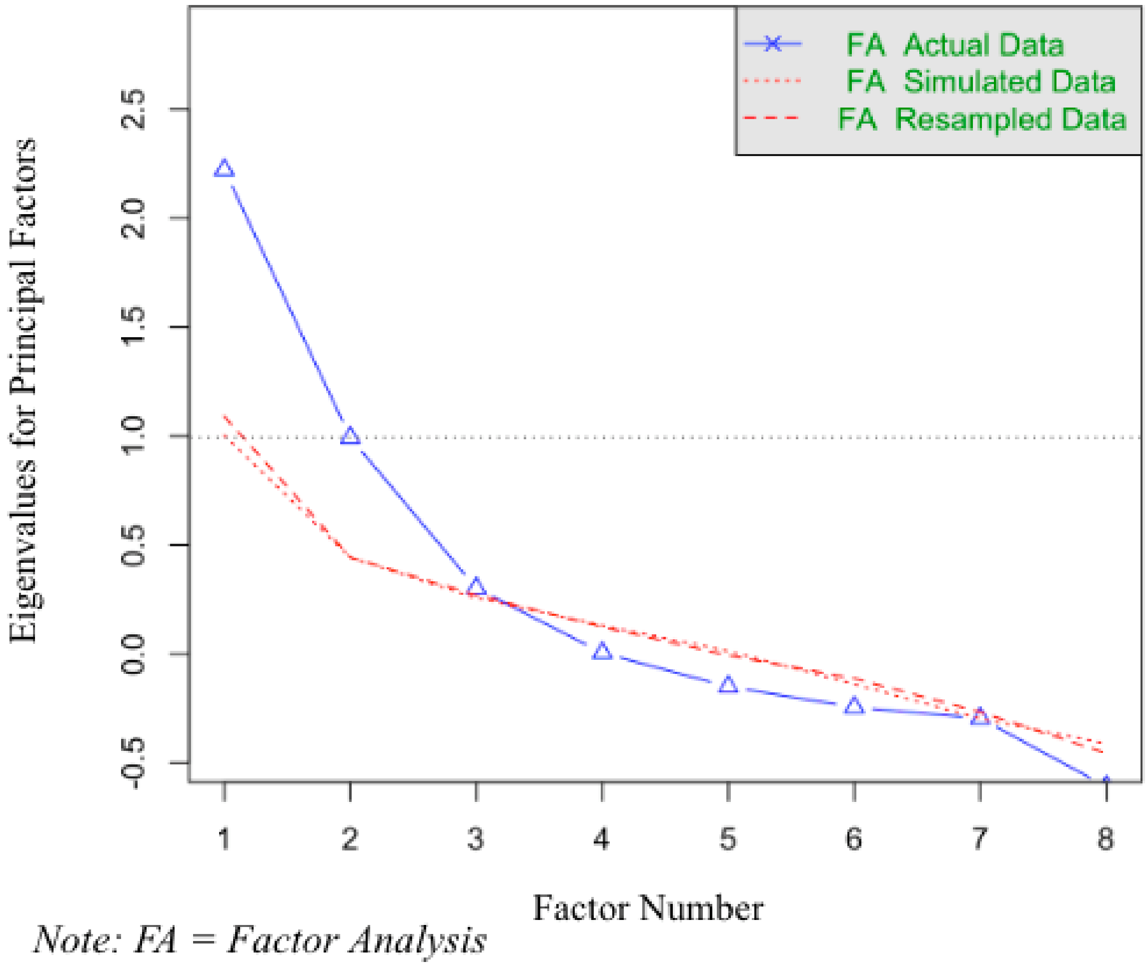


Figure 11. A scree plot suggested a two-factor structure for *Total TRAVEL scores*: all eight eye tracking and behavioral variables loaded onto the first two components and eigenvalues for both components were $\geq$1.

Figure 12. ROC Curve: Child-Directed JA Factor Score

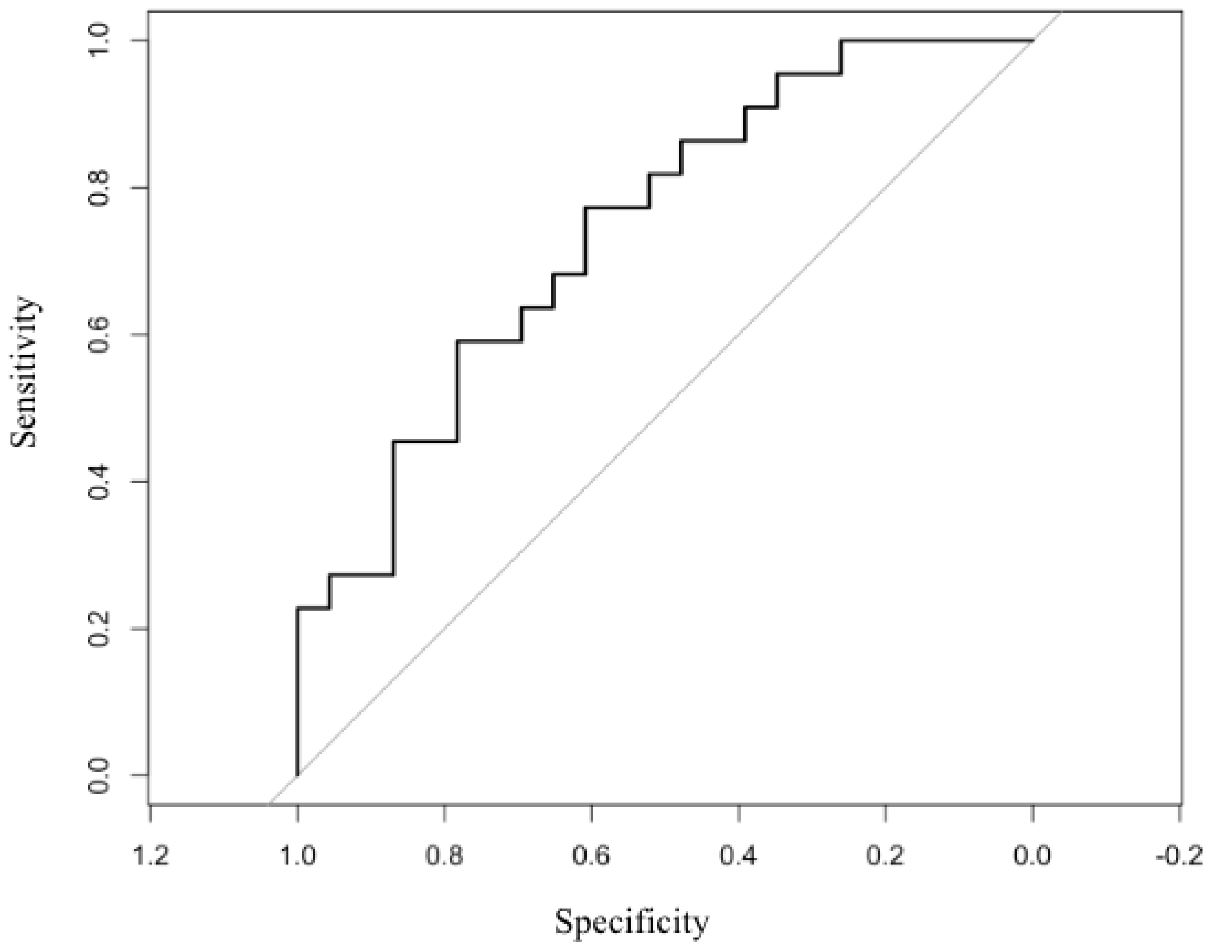


Figure 12. Receiver operator characteristic (ROC) curves with *AUC* (area under the curve) measured how accurately TRAVEL outcome measures discriminated between autistic and non-autistic groups, ranging from 0.5 to 1. An *AUC* of 0.704 for *child-directed JA* indicated that factor scores of *Eye Contact* and *Gestures* on the *PCX* (*child-directed JA*) had a 70% chance of discriminating between autistic and non-autistic children.

Figure 13. ROC Curve: Adult-Directed JA Factor Score

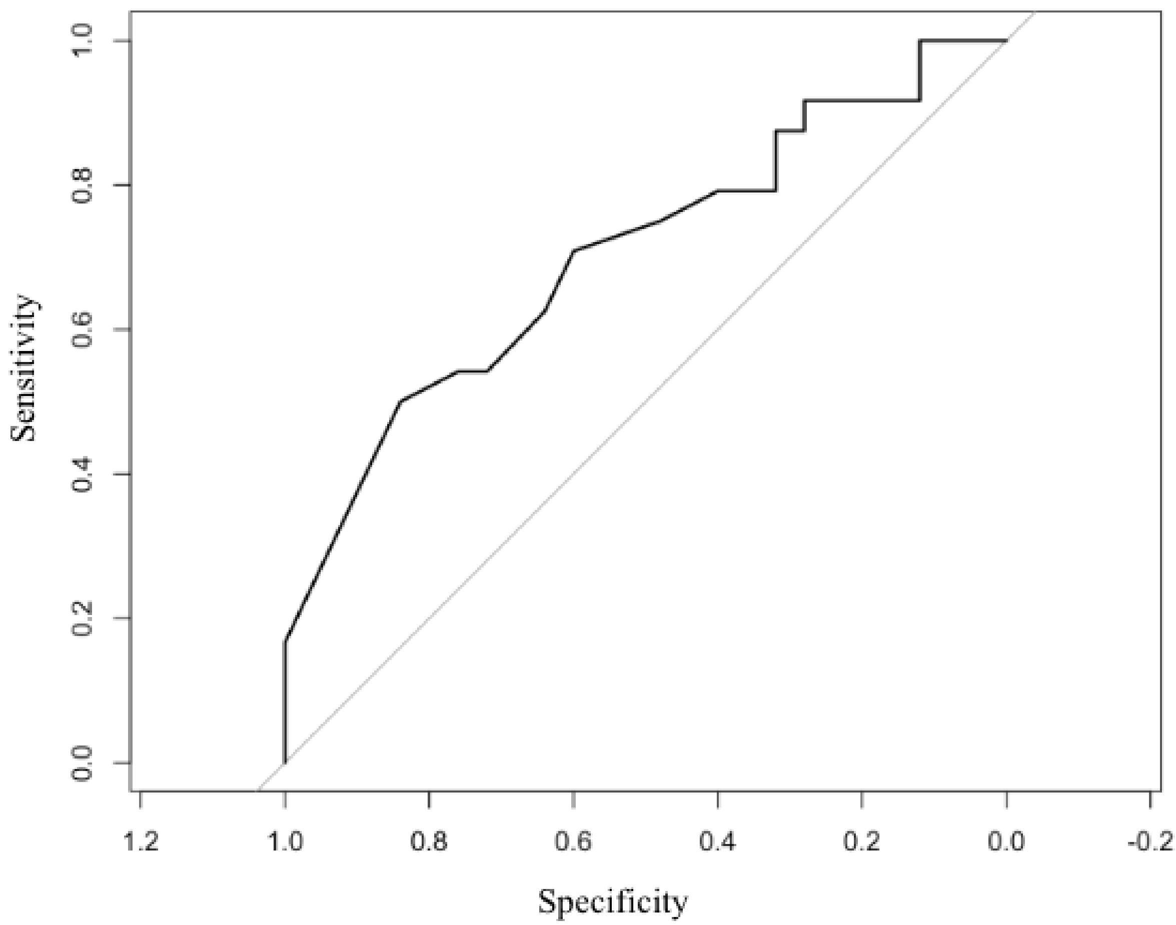


Figure 13. Receiver operator characteristic (ROC) curves with *AUC* (area under the curve) measured how accurately TRAVEL outcome measures discriminated between autistic and non-autistic groups, ranging from 0.5 to 1. An *AUC* of 0.743 for *adult-directed JA* indicated that factor scores of all eye tracking tasks and *RTN* (*adult-directed JA*) had a 74% chance of discriminating between autistic and non-autistic children.

Figure 14. Correlation Matrix: TRAVEL Outcome Variables and Child Survey Measures

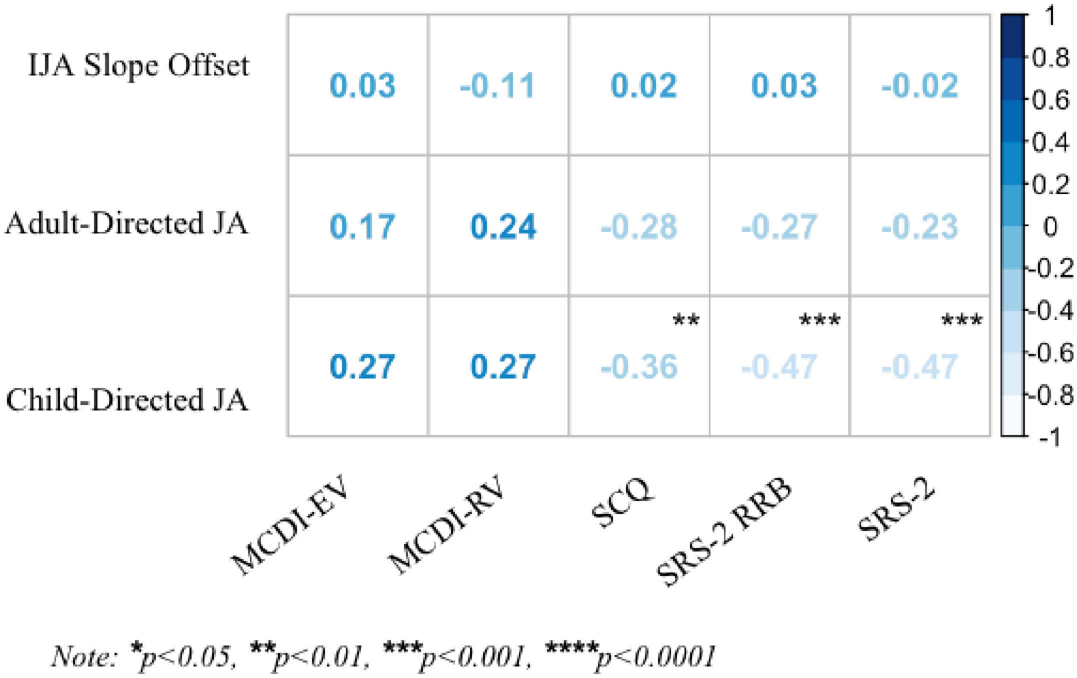


Figure 14. Pearson Partial Correlations were used for construct validity analyses (Aim 3) to measure associations between TRAVEL Outcomes and validated caregiver-report measures of joint attention (JA) (the Social Communication Questionnaire (SCQ) Total Score and the Social Responsiveness Scale (SRS-2) Total *T*-Score); of receptive (MCDI-RV) and expressive (MCDI-EV) vocabulary scores on the Mac-Arthur Bates Communicative Developmental Inventory (MCDI); and of Restricted, Repetitive Behaviors and Interests factor scores (SRS-RRBs) on the SRS-2.

REFERENCES

Acra, C. F., Bono, K. E., Mundy, P. C., & Scott, K. G. (2009). Social competence in children at risk due to prenatal cocaine exposure: Continuity over time and associations with cognitive and language abilities. *Social Development*, *18*(4), 1002–1014. https://doi.org/10.1111/j.1467-9507.2008.00519.x

Aily, J. B., Copson, J., Voinier, D., Jakiela, J., Hinman, R., Grosch, M., Noonan, C., Armellini, M., Schmitt, L., White, M., & White, D. (2023). Understanding Recruitment Yield From Social Media Advertisements and Associated Costs of a Telehealth Randomized Controlled Trial: Descriptive Study. *Journal of Medical Internet Research*, *25*, e41358. https://doi.org/10.2196/41358

Alfuraydan, M., Croxall, J., Hurt, L., Kerr, M., & Brophy, S. (2020). Use of telehealth for facilitating the diagnostic assessment of Autism Spectrum Disorder (ASD): A scoping review. *PLoS ONE*, *15*(7 July), 1–16. https://doi.org/10.1371/journal.pone.0236415

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Anderson, Colombo, & Shaddy. (2006). Visual scanning and pupillary responses in young children with Autism Spectrum Disorder. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1238–1256. https://doi.org/10.1080/13803390500376790

Astor, K., Lindskog, M., Juvrud, J., Wangchuk, Namgyel, S. C., Wangmo, T., Tshering, K., & Gredebäck, G. (2022). Maternal postpartum depression impacts infants' joint attention differentially across cultures. *Developmental Psychology*, *58*(12), 2230–2238. https://doi.org/10.1037/dev0001413

Bailey, D. B., Raspa, M., Olmsted, M., & Holiday, D. B. (2008). Co-occurring conditions associated with FMR1 gene variations: Findings from a national parent survey. *American Journal of Medical Genetics Part A*, *146A*(16), 2060–2069. https://doi.org/10.1002/ajmg.a.32439

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson, C., Rosenberg, White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L.-C., Harrington, R., Lopez, M., Fitzgerald, R. T., … Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United

States, 2014. *MMWR. Surveillance Summaries*, *67*(6), 1–23. https://doi.org/10.15585/mmwr.ss6706a1

Baird, J., Stevenson, J. C., & Williams, D. C. (2000). The Evolution of ADHD: A Disorder of Communication? *The Quarterly Review of Biology*, *75*(1), 17–35.

Bakeman, R., & Adamson, L. B. (1984). Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. *Child Development*, *55*(4), 1278–1289.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In *Joint attention: Its origins and role in development.* (pp. 131–158). Lawrence Erlbaum Associates, Inc.

Baranek, G. T. (1999). Autism During Infancy: A Retrospective Video Analysis of Sensory-Motor and Social Behaviors at 9-12 Months of Age. In *Journal of Autism and Developmental Disorders* (Vol. 29, Issue 3). https://doi.org/10.1023/A:1023080005650

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental Science*, *14*(2), 353–359. https://doi.org/10.1111/j.1467-7687.2010.00985.x

Bargiela, S., Steward, R., & Mandy, W. (2016). The Experiences of Late-diagnosed Women with Autism Spectrum Conditions: An Investigation of the Female Autism Phenotype. *Journal of Autism and Developmental Disorders*, *46*(10), 3281–3294. https://doi.org/10.1007/s10803-016-2872-8

Barnard-Brak, L., Brewer, A., Chesnut, S., Richman, D., & Schaeffer, A. M. (2016). The sensitivity and specificity of the social communication questionnaire for autism spectrum with respect to age. *Autism Research*, *9*(8), 838–845. https://doi.org/10.1002/aur.1584

Bayley, N. (1993). *The Bayley Scales of Infant Development* (2nd ed.). Psychological Corporation.

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., & Johnson, M. H. (2012). Precursors to social and communication difficulties in infants at-risk for autism: Gaze following and attentional engagement. *Journal of Autism and Developmental Disorders*, *42*(10), 2208–2218. https://doi.org/10.1007/s10803-012-1450-y

Benson-Goldberg, S., & Erickson, K. (2021). Eye-trackers, digital-libraries, and print-referencing: A single case study in CDKL5. *Research in Developmental Disabilities*, *112*, 103913. https://doi.org/10.1016/j.ridd.2021.103913

Bölte, S., Westerwald, E., Holtmann, M., Freitag, C., & Poustka, F. (2011). Autistic traits and autism spectrum disorders: The clinical validity of two measures presuming a continuum of social communication skills. *Journal of Autism and Developmental Disorders*, *41*(1), 66–72. https://doi.org/10.1007/s10803-010-1024-9

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How We Design Feasibility Studies. In *American Journal of Preventive Medicine* (Vol. 36, Issue 5, pp. 452–457). https://doi.org/10.1016/j.amepre.2009.02.002

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Bradshaw, J., Shic, F., & Chawarska, K. (2011). Brief report: Face-specific recognition deficits in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(10), 1429–1435. https://doi.org/10.1007/s10803-010-1150-4

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. In *Developmental Science* (Vol. 8, Issue 6).

Brown, T. A. (2006). Confirmatory Factor Analysis for Applied Research. In *The American Statistician* (Issue 1). The Guilford Press. https://doi.org/10.1198/tas.2008.s98

Bruner. (1974). From communication to language-a psychological perspective. *Cognition*, *3*(3), 255–287. https://doi.org/10.1016/0010-0277(74)90012-2

Cage, E., Monaco, J. Di, & Newell, V. (2018). Experiences of Autism Acceptance and Mental Health in Autistic Adults. *Journal of Autism and Developmental Disorders*, *48*, 473–484. https://doi.org/10.1007/s10803-017-3342-7

Canfield, R. L., & Kirkham, N. Z. (2001). Infant Cortical Development and the Prospective Control of Saccadic Eye Movements. *Infancy*, *2*(2), 197–211. https://doi.org/10.1207/S15327078IN0202_5

Carpenter, M., & Tomasello, M. (2000). Joint attention, cultural learning, and language acquisition: Implications for children with autism. In A. M. Wetherby & B. M. Prizant (Eds.), *Communication and language issues in autism and pervasive developmental disorder: A transactional developmental perspective* (pp. 31–54). Brookes.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 315–324. https://doi.org/10.1098/rstb.2002.1199

Charman, T. (2004). *Matching Preschool Children with Autism SpectrumDisorders and Comparison Children for Language Ability: Methodological Challenges*.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur communicative development inventory (Infant Form). *Journal of Child Language*, *30*(1), 213–236. https://doi.org/10.1017/S0305000902005482

Charman, T., & Gotham, K. (2013). Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practise. *Child and Adolescent Mental Health*, *18*(1), 52–63. https://doi.org/10.1111/j.1475-3588.2012.00664.x

Chawarska, K., Macari, S., Powell, K., Dinicola, L., & Shic, F. (2016). Enhanced Social Attention in Female Infant Siblings at Risk for Autism. In *J Am Acad Child Adolesc Psychiatry* (Vol. 55, Issue 3). www.jaacap.org

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, *74*(3), 195–203. https://doi.org/10.1016/j.biopsych.2012.11.022

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-Year-old children with Autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(12), 1663–1672. https://doi.org/10.1007/s10803-009-0803-7

Chawarska, K., & Volkmar, F. (2007). Impairments in monkey and human face recognition in 2-year-old toddlers with Autism Spectrum Disorder and developmental delay. In *Developmental Science* (Vol. 10, Issue 2, pp. 266–279). https://doi.org/10.1111/j.1467-7687.2006.00543.x

Chawarska, K., Volkmar, F., & Klin, A. (2010). Limited attentional bias for faces in toddlers with autism spectrum disorders. *Archives of General Psychiatry*, *67*(2), 178–185. https://doi.org/10.1001/archgenpsychiatry.2009.194

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. In *Research in Developmental Disabilities* (Vol. 48, pp. 79–93). Elsevier Inc. https://doi.org/10.1016/j.ridd.2015.10.011

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. In *Psychological Assessment* (Issue 4).

Congiu, S., Fadda, R., Doneddu, G., & Striano, T. (2016). Impaired representational gaze following in children with autism spectrum disorder. *Research in Developmental Disabilities*, *57*, 11–17. https://doi.org/10.1016/j.ridd.2016.06.008

Constantino, J., & Gruber, C. (2012). *Social Responsiveness Scale* (2nd ed.). Western Psychological Services.

Cooper, K., Smith, L. G. E., & Russell, A. (2017). Social identity, self-esteem, and mental health in autism. *European Journal of Social Psychology*, *47*(7), 844–854. https://doi.org/10.1002/ejsp.2297

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

Corona, L., Hine, J., Nicholson, A., Stone, C., Swanson, A., Wade, J., Wagner, L., Weitlauf, A., & Warren, Z. (2020). *TELE-ASD-PEDS: A Telemedicine-based ASD Evaluation Tool for Toddlers and Young Children*. https://vkc.vumc.org/vkc/triad/tele-asd-peds

Corona, Weitlauf, A. S., Hine, J., Berman, A., Miceli, A., Nicholson, A., Stone, C., Broderick, N., Francis, S., Juárez, A. P., Vehorn, A., Wagner, L., & Warren, Z. (2021). Parent Perceptions of Caregiver-Mediated Telemedicine Tools for Assessing Autism Risk in Toddlers. *Journal of Autism and Developmental Disorders*, *51*(2), 476–486. https://doi.org/10.1007/s10803-020-04554-9

Counts, N. (2023). *Understanding the U.S. Behavioral Health Workforce Shortage*. https://doi.org/https://doi.org/10.26099/5km6-8193

Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, *16*(2), 239–249. https://doi.org/10.1017/S0305000900010394

Dawson, G., Munson, J., Estes, A., Osterling, J., Mcpartland, J., Toth, K., Carver, L., & Abbott, R. (2002). Neurocognitive Function and Joint Attention Ability in Young Children with Autism Spectrum Disorder Versus Developmental Delay. In *Child Development* (Vol. 73, Issue 2).

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Dow, D., Day, T. N., Kutta, T. J., Nottke, C., & Wetherby, A. M. (2020). Screening for autism spectrum disorder in a naturalistic home setting using the systematic observation of red flags (SORF) at 18–24 months. *Autism Research*, *13*(1), 122–133. https://doi.org/10.1002/aur.2226

Dow, D., Guthrie, W., Stronach, S. T., & Wetherby, A. M. (2017). Psychometric analysis of the Systematic Observation of Red Flags for autism spectrum disorder in toddlers. *Autism*, *21*(3), 301–309. https://doi.org/10.1177/1362361316636760

Dow, D., Holbrook, A., Toolan, C., McDonald, N., Sterrett, K., Rosen, N., Kim, S. H., & Lord, C. (2021). The Brief Observation of Symptoms of Autism (BOSA): Development of a New Adapted Assessment Measure for Remote Telehealth Administration Through COVID-19 and Beyond. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-021-05395-w

Dunham, P. J., Dunham, F., & Curwin, A. (1993). Joint-Attentional States and Lexical Acquisition at 18 Months. *Developmental Psychology*, *29*(5), 827–831. https://doi.org/10.1037/0012-1649.29.5.827

Dykens, E. M., Lee, E., & Roof, E. (2011). Prader-Willi syndrome and autism spectrum disorders: An evolving story. *Journal of Neurodevelopmental Disorders*, *3*(3), 225–237. https://doi.org/10.1007/s11689-011-9092-5

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L. H., Espinosa, S., Campbell, K., Brotkin, S., Schaich-Borg, J., Qiu, Q., Tepper, M., Baker, J. P., Bloomfield, R. A., & Sapiro, G. (2018). Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *Npj Digital Medicine*, *1*(1). https://doi.org/10.1038/s41746-018-0024-6

Ellison, K. S., Guidry, J., Picou, P., Adenuga, P., & Davis, T. E. (2021). Telehealth and Autism Prior to and in the Age of COVID-19: A Systematic and Critical Review of the Last Decade. In *Clinical Child and Family Psychology Review* (Vol. 24, Issue 3, pp. 599–630). Springer. https://doi.org/10.1007/s10567-021-00358-0

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581–604. www.elsevier.com/locate/neubiorev

Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J., & Dawson, G. (2015). Long-Term Outcomes of Early Intervention in 6-Year-Old Children With Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(7), 580–587. https://doi.org/10.1016/j.jaac.2015.04.005

Factor, R. S., Ryan, S. M., Farley, J. P., Ollendick, T. H., & Scarpa, A. (2017). Does the Presence of Anxiety and ADHD Symptoms Add to Social Impairment in Children with Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders*, *47*(4), 1122–1134. https://doi.org/10.1007/s10803-016-3025-9

Fadus, M. C., Ginsburg, K. R., Sobowale, K., Halliday-Boykins, C. A., Bryant, B. E., Gray, K. M., & Squeglia, L. M. (2020). Unconscious Bias and the Diagnosis of Disruptive Behavior Disorders and ADHD in African American and Hispanic Youth. *Academic Psychiatry*, *44*(1), 95–102. https://doi.org/10.1007/s40596-019-01127-6

Falck-Ytter, T., Rehnberg, E., & Bölte, S. (2013). Lack of Visual Orienting to Biological Motion and Audiovisual Synchrony in 3-Year-Olds with Autism. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068816

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye Contact Detection in Humans from Birth. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Paul H. Brookes Publishing Co.

Fitzgerald, H. E. (1968). Autonomic pupillary reflex activity during early infancy and its relation to social and nonsocial visual stimuli. *Journal of Experimental Child Psychology*, *6*(3), 470–482. https://doi.org/10.1016/0022-0965(68)90127-6

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527.

Gaze Viewer. (2022). Tobii Dynavox. https://www.tobiidynavox.com/products/gaze-viewer?tab=1

George, S., Duran, N., & Norris, K. (2014). A Systematic Review of Barriers and Facilitators to Minority Research Participation Among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American Journal of Public Health*, *104*(2), e16–e31. https://doi.org/10.2105/AJPH.2013.301706

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *20*(6), 642–651. https://doi.org/10.1017/S1355617714000411

Goldberg, W. A., Jarvis, K. L., Osann, K., Laulhere, T. M., Straub, C., Thomas, E., Filipek, P., & Spence, M. A. (2005). Brief report: Early social communication behaviors in the younger siblings of children with autism. *Journal of Autism and Developmental Disorders*, *35*(5), 657–664. https://doi.org/10.1007/s10803-005-0009-6

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2006). The autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, *37*(4), 613–627. https://doi.org/10.1007/s10803-006-0280-1

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring Changes in Social Communication Behaviors: Preliminary Development of the Brief Observation of Social Communication Change (BOSCC). *Journal of Autism and Developmental Disorders*, *46*(7), 2464–2479. https://doi.org/10.1007/s10803-016-2782-9

Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*(3), 303–329. https://doi.org/10.1207/s15327906mbr4003_2

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye Contact and Face Scanning in Early Infancy. In *New Series* (Vol. 198, Issue 4319).

Hall, W. J., Chapman, M. V., Lee, K. M., Merino, Y. M., Thomas, T. W., Payne, B. K., Eng, E., Day, S. H., & Coyne-Beasley, T. (2015). Implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: A systematic review. In *American Journal of Public Health* (Vol. 105, Issue 12, pp. e60–e76). American Public Health Association Inc. https://doi.org/10.2105/AJPH.2015.302903

Harrop, C., McConachie, H., Emsley, R., Leadbitter, K., & Green, J. (2014). Restricted and repetitive behaviors in autism spectrum disorders and typical development: Cross-sectional and longitudinal comparisons. *Journal of Autism and Developmental Disorders*, *44*(5), 1207–1219. https://doi.org/10.1007/s10803-013-1986-5

Hashemi, J., Campbell, K., Carpenter, K. L. H., Harris, A., Qiu, Q., Tepper, M., Espinosa, S., Borg, J. S., Marsan, S., Calderbank, R., Baker, J. P., Egger, H. L., Dawson, G., & Sapiro, G. (2016). A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. *EAI Endorsed Transactions on Scalable Information Systems*, *3*(10), 1–5. https://doi.org/10.4108/eai.14-10-2015.2261939

Head, A. M., McGillivray, J. A., & Stokes, M. A. (2014). Gender differences in emotionality and sociability in children with autism spectrum disorders. *Molecular Autism*, *5*(1), 1–9. https://doi.org/10.1186/2040-2392-5-19

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *New Series*, *143*(3611), 1190–1192. http://www.jstor.org/stable/1712692

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, *9*(2), 131–134. https://doi.org/10.1111/1467-9280.00024

Hyde, C., Pizzano, M., McDonald, N. M., Nelson, C. A., Kasari, C., Thiele, E. A., & Jeste, S. S. (2020). A telehealth approach to improving clinical trial access for infants with tuberous sclerosis complex. *Journal of Neurodevelopmental Disorders*, *12*(1). https://doi.org/10.1186/s11689-019-9302-0

Jones, & Carr. (2004). Joint Attention in Children With Autism: Theory and Intervention. *Focus on Autism and Other Developmental Disabilities*, *19*(1), 13–26. https://doi.org/10.1177/10883576040190010301

Jones, Carr, & Klin. (2008). Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers With Autism Spectrum Disorder. *ARCH GEN PSYCHIATRY*, *65*(8).

Jones, & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Jones, W., Klaiman, C., Richardson, S., Aoki, C., Smith, C., Minjarez, M., Bernier, R., Pedapati, E., Bishop, S., Ence, W., Wainer, A., Moriuchi, J., Tay, S.-W., & Klin, A. (2023a). Eye-Tracking–Based Measurement of Social Visual Engagement Compared With Expert Clinical Diagnosis of Autism. *JAMA*, *330*(9), 854. https://doi.org/10.1001/jama.2023.13295

Jones, W., Klaiman, C., Richardson, S., Lambha, M., Reid, M., Hamner, T., Beacham, C., Lewis, P., Paredes, J., Edwards, L., Marrus, N., Constantino, J. N., Shultz, S., & Klin, A. (2023b). Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism. *JAMA Network Open*, *6*(9), e2330145. https://doi.org/10.1001/jamanetworkopen.2023.30145

Kasari, C., Freeman, S., Mundy, P., & Sigman, M. D. (1995). Attention regulation by children with Down syndrome: coordinated joint attention and social referencing looks. *American Journal of Mental Retardation : AJMR*, *100*(2), 128–136.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(6), 611–620. https://doi.org/10.1111/j.1469-7610.2005.01567.x

Kelleher, B. L., Halligan, T., Witthuhn, N., Neo, W. S., Hamrick, L., & Abbeduto, L. (2020). Bringing the Laboratory Home: PANDABox Telehealth-Based Assessment of Neurodevelopmental Risk in Children. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01634

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Defining and Quantifying the Social Phenotype in Autism. *American Journal of Psychiatry*, *159*(6), 895–908. https://doi.org/10.1176/appi.ajp.159.6.895

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*. https://doi.org/doi:10.1038/nature07868

Lecavalier, L., Bodfish, J., Harrop, C., Whitten, A., Jones, D., Pritchett, J., Faldowski, R., & Boyd, B. (2020). Development of the Behavioral Inflexibility Scale for Children with Autism Spectrum Disorder and Other Developmental Disabilities. *Autism Research*, *13*(3), 489–499. https://doi.org/10.1002/aur.2257

Lee, C. M., Green Snyder, L., Carpenter, L. A., Harris, J., Kanne, S., Taylor, C. M., Sarver, D. E., Stephenson, K. G., Shulman, L. H., Wodka, E. L., Esler, A., & SPARK Consortium. (2023). Agreement of parent-reported cognitive level with standardized measures among children with autism spectrum disorder. *Autism Research : Official Journal of the International Society for Autism Research*. https://doi.org/10.1002/aur.2934

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and Cues: Gaze-following in Children with Autism. *Journal of Child Psychology and Psychiatry*, *39*(7), S0021963098003035. https://doi.org/10.1017/S0021963098003035

Lerwick, J. L. (2016). Minimizing pediatric healthcare-induced anxiety and trauma. *World Journal of Clinical Pediatrics*, *5*(2), 143. https://doi.org/10.5409/wjcp.v5.i2.143

Lin, T.-Y., Huang, C.-C., Chiang, C.-H., Kuo, P.-L., & Lin, S.-J. (2003). Nonverbal communication skills in children with Prader-Willi syndrome (PWS). *Tzu Chi Medical Journal*, *15*(5), 311–320.

Lord, C., Risi, S., Lambrecht, L., Cook, Jr. , E. H., Lenventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). *The Autism Diagnostic Observation Schedule–Generic:A Standard Measure of Social and Communication DeficitsAssociated with the Spectrum of Autism*.

Loveland, K. A., & Landry, S. H. (1986). Joint Attention and Language in Autism and Developmental Language Delay I. In *Journal of Autism and Developmental Disorders* (Vol. 16, Issue 3).

Maenner, M. J., Shaw, K. A., Bakian, A. v., Bilder, D. A., Durkin, M. S., Esler, A., Furnier, S. M., Hallas, L., Hall-Lande, J., Hudson, A., Hughes, M. M., Patrick, M., Pierce, K., Poynter, J. N., Salinas, A., Shenouda, J., Vehorn, A., Warren, Z., Constantino, J. N., … Cogswell, M. E. (2021). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR. Surveillance Summaries*, *70*(11), 1–16. https://doi.org/10.15585/mmwr.ss7011a1

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. v., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., … Shaw, K. A. (2023). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR. Surveillance Summaries, 72*(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. v., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., … Shaw, K. A. (2023). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR. Surveillance Summaries*, *72*(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Mandell, D. S., Ittenbach, R. F., Levy, S. E., & Pinto-Martin, J. A. (2007). Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*(9), 1795–1802. https://doi.org/10.1007/s10803-006-0314-8

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., Giarelli, E., Morrier, M. J., Nicholas, J. S., Pinto-Martin, J. A., Shattuck, P. T., Thomas, K. C., Yeargin-Allsopp, M., & Kirby, R. S. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, *99*(3), 493–498. https://doi.org/10.2105/AJPH.2007.131243

Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*(3), 302–315. https://doi.org/10.1111/1467-9507.00127

Maurer, D., & Salapatek, P. (1976). Developmental Changes in the Scanning of Faces by Young Infants. *Child Development*, *47*(2), 523. https://doi.org/10.2307/1128813

Moody, E. J., Reyes, N., Ledbetter, C., Wiggins, L., DiGuiseppi, C., Alexander, A., Jackson, S., Lee, L. C., Levy, S. E., & Rosenberg, S. A. (2017). Screening for Autism with the SRS and SCQ: Variations across Demographic, Developmental and Behavioral Factors in Preschool Children. *Journal of Autism and Developmental Disorders*, *47*(11), 3550–3561. https://doi.org/10.1007/s10803-017-3255-5

Morales, M., Mundy, P., Crowson, M. M., Neal, A. R., & Delgado, C. E. F. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. *International Journal of Behavioral Development*, *29*(3), 259–263. https://doi.org/10.1177/01650250444000432

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to Joint Attention Across the 6- Through 24-Month Age Period and Early Language Acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10.1016/S0193-3973(99)00040-4

Mottron, L., Mineau, S., Martel, G., Bernier, C. S.-C., Berthiaume, C., Dawson, M., Lemay, M., Palardy, S., Charman, T., & Faubert, J. (2007). Lateral glances toward moving stimuli among young children with autism: Early regulation of locally oriented perception? *Development and Psychopathology*, *19*(01). https://doi.org/10.1017/S0954579407070022

Mundy, P. (2003). *Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system*.

Mundy, P. (2013). A neural networks, information-processing model of joint attention and social-cognitive development. In *Zelazo, Philip David [Ed]* (Vol. 2, Issue October). https://doi.org/10.1093/oxfordhb/9780199958474.013.0010

Mundy, P. (2017). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Mundy, P., & Crowson, M. (1997). Joint attention and early social communication: Implications for research on inervention with autism. *Journal of Autism and Developmental Disorders*, *27*(6), 653–676. https://doi.org/10.1023/A:1025802832021

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). Manual for early social communication scales (ESCS). *Coral Gables, FL: ...*, *305*. http://www.ucdmc.ucdavis.edu/mindinstitute/ourteam/faculty_staff/escs.pdf

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8–9), 985–997. https://doi.org/10.1016/j.neunet.2010.08.009

Mundy, P., Kim, K., McIntyre, N., Lerro, L., & Jarrold, W. (2016). Brief Report: Joint Attention and Information Processing in Children with Higher Functioning Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *46*(7), 2555–2560. https://doi.org/10.1007/s10803-016-2785-6

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Mundy, P., Novotny, S., Swain-Lerro, L., McIntyre, N., Zajic, M., & Oswald, T. (2017). Joint-Attention and the Social Phenotype of School-Aged Children with ASD. *Journal of Autism and Developmental Disorders*, *47*(5), 1423–1435. https://doi.org/10.1007/s10803-017-3061-0

Mundy, P., & Sigman, M. (2015). Joint Attention, Social Competence, and Developmental Psychopathology. In *Developmental Psychopathology: Second Edition* (Vol. 1, Issue March). https://doi.org/10.1002/9780470939383.ch9

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128. https://doi.org/10.1007/BF02206861

Mundy, P., Sigman, M., & Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology*, *6*(3), 389–401. https://doi.org/10.1017/S0954579400006003

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry*, *27*(5), 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, *2*(1), 2–21. https://doi.org/10.1002/aur.61

Mundy, P., & Vaughan Van Hecke, A. (2008). Neural Systems, Gaze Following, and the Development of Joint Attention. *Handbook of Developmental Cognitive Neuroscience*, *May*. https://doi.org/10.7551/mitpress/7437.003.0059

Murias, M., Major, S., Davlantis, K., Franz, L., Harris, A., Rardin, B., Sabatos-DeVito, M., & Dawson, G. (2018). Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials. *Autism Research*, *11*(1), 166–174. https://doi.org/10.1002/aur.1894

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology*, *20*(1), 79–97. https://doi.org/10.1017/S0954579408000047

Nazneen, N., Rozga, A., Smith, C. J., Oberleitner, R., Abowd, G. D., & Arriaga, R. I. (2015). A Novel System for Supporting Autism Diagnosis Using Home Videos: Iterative Development and Evaluation of System Design. *JMIR MHealth and UHealth*, *3*(2), 1–14. https://doi.org/10.2196/mhealth.4393

Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research*, *13*(6), 1–15. https://doi.org/10.16910/jemr.13.6.1

NLSY Child Handbook: A Guide & Resource for the NLSY 1986 Child Data (PDF). (1989).

Nowell, S. W., Watson, L. R., Faldowski, R. A., & Baranek, G. T. (2018). An Initial Psychometric Evaluation of the Joint Attention Protocol. *Journal of Autism and Developmental Disorders*, *48*(6), 1932–1944. https://doi.org/10.1007/s10803-017-3458-9

Nyström, P., Thorup, E., Bölte, S., & Falck-Ytter, T. (2019). Joint Attention in Infancy and the Emergence of Autism. *Biological Psychiatry*, *86*(8), 631–638. https://doi.org/10.1016/j.biopsych.2019.05.006

Osterling, J., & Dawson, G. (1994). Early Recognition of Children with Autism: A Study of First Birthday Home Videotapes 1. In *Journal of Autism and Developmental Disorders* (Vol. 24, Issue 3).

Ozonoff, S., Macari, S., Young, G. S., Goldring, S., Thompson, M., & Rogers, S. J. (2008). Atypical object exploration at 12 months of age is associated with autism in a prospective sample. *Autism*, *12*(5), 457–472. https://doi.org/10.1177/1362361308096402

Parsons, J. P., Bedford, R., Jones, E. J., Charman, T., Johnson, M. H., & Gliga, T. (2019). Gaze following and attention to objects in infants at familial risk for ASD. *Frontiers in Psychology*, *10*(JULY). https://doi.org/10.3389/fpsyg.2019.01799

Pelphrey, K., Sasson, N., Reznick, J., Gregory, P., Goldman, B., & Piven, J. (2002). Visual Scanning of Faces in Autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–261. https://doi.org/https://doi-org.libproxy.lib.unc.edu/10.1023/A:1016374617369

Pelphrey, Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, *128*(5), 1038–1048. https://doi.org/10.1093/brain/awh404

Peterson, J., & Moore, K. (1987). *Motor and social development in infancy: some results from a national survey. No. 87-03.* . Child Trends, Inc.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1377), 1915–1927. https://doi.org/10.1098/rstb.1998.0344

Ravindran, V., Osgood, M., Sazawal, V., Solorzano, R., & Turnacioglu, S. (2019). Virtual reality support for joint attention using the floreo joint attention module: Usability and feasibility pilot study. *JMIR Pediatrics and Parenting*, *2*(2). https://doi.org/10.2196/14429

Richerson, J. E., Simon, G. R., Abularrage, J. J., Boudreau, A. D. A., Baker, C. N., Barden, G. A., Brown, O. “Skip” W., Hackell, J. M., Hardin, A. P., Meade, K. E., Moore, S. B., Skatud, A., Curry, E. S., Duncan, P. M., Hagan, J. F., Shaw, J. S., & Swanson, J. T. (2017). 2017 Recommendations for Preventive Pediatric Health Care. *Pediatrics*, *139*(4). https://doi.org/10.1542/peds.2017-0254

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire: Manual*. Western Psychological Services.

Salley, B., Sheinkopf, S. J., Rebecca Neal-Beevers, A., Tenenbaum, E. J., Miller-Loncar, C. L., Tronick, E., Lagasse, L. L., Shankaran, S., Bada, H., Bauer, C., Whitaker, T., Hammond, J., & Lester, B. M. (2016). Infants’ early visual attention and social engagement as developmental precursors to joint attention. *Developmental Psychology*, *52*(11), 1721–1731. https://doi.org/10.1037/dev0000205

Sanefuji, W., Wada, K., Yamamoto, T., Mohri, I., & Taniike, M. (2014). Development of preference for conspecific faces in human infants. *Developmental Psychology*, *50*(4), 979–985. https://doi.org/10.1037/a0035205

Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. *Journal of Visualized Experiments : JoVE*, *61*. https://doi.org/10.3791/3675

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Brief report: Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders*, *41*(2), 242–247. https://doi.org/10.1007/s10803-010-1038-3

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S. L., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, *1*(1), 31–42. https://doi.org/10.1002/aur.4

Schertz, H. H., & Odom, S. L. (2007). Promoting joint attention in toddlers with autism: A parent-mediated developmental model. *Journal of Autism and Developmental Disorders*, *37*(8), 1562–1575. https://doi.org/10.1007/s10803-006-0290-z

Schiltz, H. K., Mcvey, A. J., Wozniak, B. D., Haendel, A. D., Stanley, R., Arias, A., Gordon, N., & Van Hecke, A. V. (2021). The role of loneliness as a mediator between autism features and mental health among autistic young adults. *Autism*, *25*(2), 545–555. https://doi.org/10.1177/1362361320967789

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender Differences in the Social Motivation and Friendship Experiences of Autistic and Non-autistic Adolescents. *Journal of Autism and Developmental Disorders*, *46*(4), 1297–1306. https://doi.org/10.1007/s10803-015-2669-1

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., Lee, L. C., Rice, C., Giarelli, E., Kirby, R., Baio, J., Pinto-Martin, J., & Cuniff, C. (2009). Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(5), 474–483. https://doi.org/10.1097/CHI.0b013e31819b3848

Sheinkopf, S. J., Mundy, P., Claussen, A. H., & Willoughby, J. (2004). Infant joint attention skill and preschool behavioral outcomes in at-risk children. *Development and Psychopathology*, *16*(2), 273–291. https://doi.org/10.1017/S0954579404044517

Shi, L., Zhou, Y., Ou, J., Gong, J., Wang, S., Cui, X., Lyu, H., Zhao, J., & Luo, X. (2015). Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0122280

Sigman, M., & Ruskin, E. (1999). Continuity and Change in the Social Competence of Children With Autism, Down Syndrome, and Developmental Delays. *Monographs of the Society for Research in Child Development*, *64*(1, Serial No. 256).

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 809–813. https://doi.org/10.1073/pnas.0707021105

Smith, C. J., Rozga, A., Matthews, N., Oberleitner, R., Nazneen, N., & Abowd, G. (2017). Investigating the accuracy of a novel telehealth diagnostic approach for autism spectrum disorder. *Psychological Assessment*, *29*(3), 245–252. https://doi.org/10.1037/pas0000317

Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems. *Science*, *240*(4857), 1285–1293.

Tenenbaum, E. J., Major, S., Carpenter, K. L. H., Howard, J., Murias, M., & Dawson, G. (2021). Distance from Typical Scan Path When Viewing Complex Stimuli in Children with Autism Spectrum Disorder and its Association with Behavior. *Journal of Autism and Developmental Disorders*, *51*(10), 3492–3505. https://doi.org/10.1007/s10803-020-04812-w

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., & Falck-Ytter, T. (2016). Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular Autism*, *7*(1). https://doi.org/10.1186/s13229-016-0069-9

Thorup, E., Nyström, P., Gredebäck, G., Bölte, S., & Falck-Ytter, T. (2018). Reduced Alternating Gaze During Social Interaction in Infancy is Associated with Elevated Symptoms of Autism in Toddlerhood. *Journal of Abnormal Child Psychology*, *46*(7), 1547–1561. https://doi.org/10.1007/s10802-017-0388-0

Tobii Pro AB. (2014). *Tobii Pro Lab* (Version 1.207). Tobii Pro AB.

Tomasello, M., & Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development, 70*.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*(12), 197–211. https://doi.org/10.1177/014272378300401202

Toth, K., Munson, J., Meltzoff, A. N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders*, *36*(8), 993–1005. https://doi.org/10.1007/s10803-006-0137-7

Vaughan Van Hecke, A., Mundy, P., Block, J. J., Delgado, C. E. F., Parlade, M. V., Pomares, Y. B., & Hobson, J. A. (2012). Infant responding to joint attention, executive processes, and self-regulation in preschool children. *Infant Behavior and Development*, *35*(2), 303–311. https://doi.org/10.1016/j.infbeh.2011.12.001

Vaughan Van Hecke, A., Mundy, P. C., Acra, C. F., Block, J. J., Delgado, C. E. F., Parlade, M. V., Meyer, J. A., Neal, A. R., & Pomares, Y. B. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development*, *78*(1), 53–69. https://doi.org/10.1111/j.1467-8624.2007.00985.x

Vinen, Z., Clark, M., & Dissanayake, C. (2022). Social and Behavioural Outcomes of School Aged Autistic Children Who Received Community-Based Early Interventions. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05477-3

Wagner, L., Corona, L. L., Weitlauf, A. S., Marsh, K. L., Berman, A. F., Broderick, N. A., Francis, S., Hine, J., Nicholson, A., Stone, C., & Warren, Z. (2021). Use of the TELE-ASD-PEDS for Autism Evaluations in Response to COVID-19: Preliminary Outcomes and Clinician Acceptability. *Journal of Autism and Developmental Disorders*, *51*(9), 3063–3072. https://doi.org/10.1007/s10803-020-04767-y

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*, *22*(1), 25–39. https://doi.org/10.1044/1058-0360(2012/11-0145)

Watt, N., Wetherby, A., & Shumway, S. (2006). Prelinguistic predictors of language outcome at 3 years of age. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1224–1237. https://doi.org/10.1044/1092-4388(2006/088)

Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L. A., Brawley, O. W., Gross, C. P., & Emanuel, E. (2006). Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine*, *3*(2), 0201–0210. https://doi.org/10.1371/journal.pmed.0030019

Wetherby, A. M., & Prutting, C. A. (1984). Profiles of Communicative and Cognitive-Social Abilities in Autistic Children. *Journal of Speech, Language, and Hearing Research*, *27*(3), 364–377. https://doi.org/10.1044/jshr.2703.364

Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders*, *37*(5), 960–975. https://doi.org/10.1007/s10803-006-0237-4

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

Wood de Wilde, H., Kojovic, N., Robertson, C., Karr, C., Akman, L., Caccia, F., Costes, A., Etienne, M., Franchini, M., Gentaz, E., & Schaer, M. (2023). Remote Intensive Intervention for Young Children on the Autism Spectrum During COVID-19: the Experience of

Caregivers and Service Providers. *Advances in Neurodevelopmental Disorders*. https://doi.org/10.1007/s41252-023-00339-0

Wright, C. A., Kaiser, A. P., Reikowsky, D. I., & Roberts, M. Y. (2013). Effects of a naturalistic sign intervention on expressive language of toddlers with down syndrome. *Journal of Speech, Language, and Hearing Research*, *56*(3), 994–1008. https://doi.org/10.1044/1092-4388(2012/12-0060)

Yu, C., & Smith, L. B. (2016). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., Moretti, Å. H., Holm, A., Jifält, I., Kosieradzki, R., Linder, J., Nordin, V., Olafsdottir, K., Poltrago, L., & Bölte, S. (2016). The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *European Child and Adolescent Psychiatry*, *25*(7), 769–780. https://doi.org/10.1007/s00787-015-0793-2